Introduction

You have now embarked on a journey to study physics. I hope I have some way assisted you in that journey.

I have used OpenStax.com College Physics 2e develop these materials. I have used the chapters of College Physics 2e under the following license:

License:

by OpenStax is licensed under Creative Commons Attribution License v4.0

Since this is specifically written for my college students who are taking an accelerated course in physics, I have shortened the OpenStax.com college physics book to match the needs. I have done my best to not include unnecessary descriptions within this textbook, and also other information not required for this level of learning. Nevertheless, in some instances, I have included some material above this level to help you understand the fundamentals better.

I hope to continuously improve this OER textbook in the future by adding different aspects to solving problems.

Chapter 0

0.1 Introduction

This chapter includes all important aspects including the mathematical formulae useful to be successful in this physics course.

Physics becomes interesting when you relate your learning to daily activities. Therefore, the students are encouraged to reflect after each chapter, how, what, where, when the principles learnt in each chapter can be used in daily activities.

As explained in my published educational models, it is important that students do their best to develop the following six key skills while following this course. When you improve these skills while studying physics, you will reap the benefits after joining the workforce upon completion of your degree.

0.2 The Scientific Method

The scientific method is the single most important aspect a teacher should teach his/her students. Not only can the scientific method be applied to scientific experiments, but also to our daily activities, decision-making, and solving general problems. In our teaching, we should also attempt to improve the following activities within students through the subjects we teach. These key skills designated by the Department for Education in the United Kingdom are important aspects of the scientific method.

- To improve communication skills
- To improve number skills
- To improve IT skills
- To learn to work with others
- To improve own learning and performance
- To solve problems

The general steps involved in scientific method are as follows:

1. Name the problem or question.

2. Form an educated guess (hypothesis) of the cause of the problem and make predictions based upon the hypothesis.

3. Test your hypothesis by doing an experiment or study with proper controls.

4. Check and interpret your results

5. Report your results.

0.3 Solving Problems in Physics

To solve problems, you may want to implement the GUESS, or when appropriate, the GUPPESS method as follows:

Givens	**G**ivens
Unknowns	**U**nknowns
Equation	**P**rinciple
Substitution	**P**icture
Solution	**S**ubstitution
	Solution

It is also a good idea to incorporate these steps into any problem that your teacher needs you to evaluate. It makes working through your thought processes easier.

e.g:

A car accelerates at a rate of 0.60 m/s^2. How long does it take the car to go from 55 mi/hr to 60 mi/hr?

GIVENS: $a = 0.60$ m/s^2; $v_0 = 55$ mi/hr; v or $v_f = 60$ mi/hr

UNKNOWNS: t=? seconds

PRINCIPLE: You could use either of at least two equations to find the time when given acceleration and beginning and ending velocities: $a = \frac{\Delta v}{t}$ or $v = v_0 + at$

Part of the problem is that you need to get from mi/hr to m/s – and for that you use dimensional analysis

PICTURE: Probably not that helpful in this instance, but you can draw one.

EQUATION, SUBSTITUTION, and SOLUTION: $v = v_0 + at$

$$60 \text{ mi/hr} = 55 \text{ mi/hr} + (0.60 \text{ m/s}^2)t$$

$$60 \text{ mi/hr} - 55 \text{ mi/hr} = (0.60 \text{ m/s}^2)t$$

$$5 \text{ mi/hr} = (0.60 \text{ m/s}^2)t$$

Here is where you may want to use dimensional analysis in a side bar to convert 5 mi/hr into m/s^2.

$$\frac{5\text{mi}}{1\text{hr}} \times \frac{1609\text{m}}{1\text{mi}} \times \frac{1\text{hr}}{3600\text{s}} = \frac{2.2\text{m}}{1\text{s}} = 2.2\text{m/s}$$

Now substitute 2.2 m/s into the previous equation where you had 5 mi/hr.

$$5 \text{ mi/hr} = (0.60 \text{ m/s}^2)t$$

$$2.2 \text{ m/s} = (0.60 \text{ m/s}^2)t$$

$$\frac{2.2\text{m/s}}{(0.60\text{m/s}^2)} = t$$

NOTE: Follow the units! -> $\frac{2.2\text{m}}{1\text{s}} \times \frac{1\text{s}^2}{0.60\text{m}}$

3.7 s = t

0.4 Some Mathematical formulae

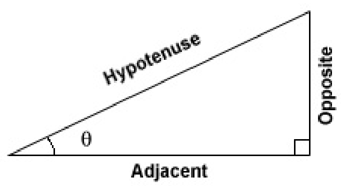
Trigonometry

Trigonometry will become important when you study vectors.

$$\sin\theta = \frac{opposite}{hypotenuse}$$

$$\tan\theta = \frac{opposite}{adjacent}$$

$$\cos\theta = \frac{adjacent}{hypotenuse}$$



.

Error Percentage Calculations

$$\% \, Change = \frac{change}{original} \; 100\,\%$$

$$\% \, Change = \frac{new - original}{original} \, 100\,\%$$

Quadratic Formula

$$ax^2 + bx + c = 0$$

x is the variable and *a*, *b*, and *c* are constants

The x value is given by the **quadratic formula**:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Calculus Formulae

Trigonometric Formulae

1. $\sin^2\theta + \cos^2\theta = 1$
2. $1 + \tan^2\theta = \sec^2\theta$
3. $1 + \cot^2\theta = \csc^2\theta$
4. $\sin(-\theta) = -\sin\theta$
5. $\cos(-\theta) = \cos\theta$
6. $\tan(-\theta) = -\tan\theta$

7. $\sin(A+B)=\sin A\cos B+\sin B\cos A$
8. $\sin(A-B)=\sin A\cos B-\sin B\cos A$
9. $\cos(A+B)=\cos A\cos B-\sin A\sin B$
10. $\cos(A-B)=\cos A\cos B+\sin A\sin B$
11. $\sin 2\theta = 2\sin\theta\cos\theta$
12. $\cos 2\theta = \cos^2\theta-\sin^2\theta = 2\cos^2\theta-1=1-2\sin^2\theta$
13. $\tan\theta = \dfrac{\sin\theta}{\cos\theta}=\dfrac{1}{\cot\theta}$
14. $\cot\theta = \dfrac{\cos\theta}{\sin\theta}=\dfrac{1}{\tan\theta}$
15. $\sec\theta = \dfrac{1}{\cos\theta}$
16. $\csc\theta = \dfrac{1}{\sin\theta}$
17. $\cos(\dfrac{\pi}{2}-\theta)=\sin\theta$
18. $\sin(\dfrac{\pi}{2}-\theta)=\cos\theta$

Differentiation Formulas

1. $\dfrac{d}{dx}(x^n)=nx^{n-1}$
2. $\dfrac{d}{dx}(fg)=fg'+gf'$
3. $\dfrac{d}{dx}(\dfrac{f}{g})=\dfrac{gf'-fg'}{g^2}$
4. $\dfrac{d}{dx}f(g(x))=f'(g(x))g'(x)$
5. $\dfrac{d}{dx}(\sin x)=\cos x$
6. $\dfrac{d}{dx}(\cos x)=-\sin x$
7. $\dfrac{d}{dx}(\tan x)=\sec^2 x$
8. $\dfrac{d}{dx}(\cot x)=-\csc^2 x$
9. $\dfrac{d}{dx}(\sec x)=\sec x\tan x$

10. $\frac{d}{dx}(\csc x) = -\csc x \cot x$

11. $\frac{d}{dx}(e^x) = e^x$

12. $\frac{d}{dx}(a^x) = a^x \ln a$

13. $\frac{d}{dx}(\ln x) = \frac{1}{x}$

14. $\frac{d}{dx}(Arc \sin x) = \frac{1}{\sqrt{1-x^2}}$

15. $\frac{d}{dx}(Arc \tan x) = \frac{1}{1+x^2}$

16. $\frac{d}{dx}(Arc \sec x) = \frac{1}{|x|\sqrt{x^2-1}}$

17. $\frac{dy}{dx} = \frac{dy}{dx} \times \frac{du}{dx}$ *Chain Rule*

Integration Formulae

1. $\int a\,dx = ax + C$
2. $\int x^n\,dx = \frac{x^{n+1}}{n+1} + C, \quad n \neq -1$
3. $\int \frac{1}{x} dx = \ln|x| + C$
4. $\int e^x\,dx = e^x + C$
5. $\int a^x dx = \frac{a^x}{\ln a} + C$
6. $\int \ln x\,dx = x \ln x - x + C$
7. $\int \sin x\,dx = -\cos x + C$
8. $\int \cos x\,dx = \sin x + C$
9. $\int \tan x\,dx = \ln|\sec x| + C \text{ or } -\ln|\cos x| + C$
10. $\int \cot x\,dx = \ln|\sin x| + C$
11. $\int \sec x\,dx = \ln|\sec x + \tan x| + C$
12. $\int \csc x\,dx = \ln|\csc x - \cot x| + C$
13. $\int \sec^2 x\,dx = \tan x + C$
14. $\int \sec x \tan x\,dx = \sec x + C$

15. $\int \csc^2 x\,dx = -\cot x + C$

16. $\int \csc x \cot x\,dx = -\csc x + C$

17. $\int \tan^2 x\,dx = \tan x - x + C$

18. $\int \frac{dx}{a^2+x^2} = \frac{1}{a} Arc\tan\left(\frac{x}{a}\right) + C$

19. $\int \frac{dx}{\sqrt{a^2-x^2}} = Arc\sin\left(\frac{x}{a}\right) + C$

20. $\int \frac{dx}{x\sqrt{x^2-a^2}} = \frac{1}{a} Arc\sec\frac{|x|}{a} + C = \frac{1}{a} Arc\cos\left|\frac{a}{x}\right| + C$

Formulae and Theorems

Definition of Limit: Let f be a function defined on an open interval containing c (except possibly at c) and let L be a real number. Then $\lim_{x \to a} f(x) = L$ means that for each $\varepsilon > 0$ there exists a $\delta > 0$ such that $|f(x) - L| < \varepsilon$ whenever $0 < |x-c| < \delta$.

1b. A function $y = f(x)$ is continuous at $x = a$ if

i). *f(a)* exists

ii). $\lim_{x \to a} f(x)$ exists

iii). $\lim_{x \to a} = f(a)$

Even and Odd Functions

1. A function $y = f(x)$ is even if $f(-x) = f(x)$ for every x in the function's domain.
   Every even function is symmetric about the y-axis.

0. A function $y = f(x)$ is odd if $f(-x) = -f(x)$ for every x in the function's domain.
   Every odd function is symmetric about the origin.

Periodicity

A function $f(x)$ is periodic with period $p\ (p > 0)$ if $f(x+p) = f(x)$ for every value of x.

Note: The period of the function $y = A\sin(Bx + C)$ or $y = A\cos(Bx + C)$ is $\frac{2\pi}{|B|}$.

The amplitude is $|A|$. The period of $y = \tan x$ is π.

Intermediate-Value Theorem

A function $y = f(x)$ that is continuous on a closed interval $[a,b]$ takes on every value between $f(a)$ and $f(b)$.

Note: If f is continuous on $[a,b]$ and $f(a)$ and $f(b)$ differ in sign, then the equation $f(x) = 0$ has at least one solution in the open interval (a,b).

Limits of Rational Functions as $x \to \pm\infty$

i). $\lim\limits_{x \to \pm\infty} \frac{f(x)}{g(x)} = 0$ if the degree of $f(x) <$ the degree of $g(x)$

Example: $\lim\limits_{x \to \infty} \frac{x^2 - 2x}{x^3 + 3} = 0$

ii). $\lim\limits_{x \to \pm\infty} \frac{f(x)}{g(x)}$ is infinite if the degrees of $f(x) >$ the degree of $g(x)$

Example: $\lim\limits_{x \to \infty} \frac{x^3 + 2x}{x^2 - 8} = \infty$

iii). $\lim\limits_{x \to \pm\infty} \frac{f(x)}{g(x)}$ is finite if the degree of $f(x) =$ the degree of $g(x)$

Example: $\lim\limits_{x \to \infty} \frac{2x^2 - 3x + 2}{10x - 5x^2} = -\frac{2}{5}$

Horizontal and Vertical Asymptotes

1. A line $y = b$ is a horizontal asymptote of the graph $y = f(x)$ if either $\lim\limits_{x \to \infty} f(x) = b$ or $\lim\limits_{x \to -\infty} f(x) = b$.

0. A line $x = a$ is a vertical asymptote of the graph $y = f(x)$ if either $\lim_{x \to a^+} f(x) = \pm\infty$ or $\lim_{x \to a^-} = \pm\infty$.

Average and Instantaneous Rate of Change

i). *Average Rate of Change*: If (x_0, y_0) and (x_1, y_1) are points on the graph of $y = f(x)$, then the average rate of change of y with respect to x over the interval $[x_0, x_1]$ is $\dfrac{f(x_1) - f(x_0)}{x_1 - x_0} = \dfrac{y_1 - y_0}{x_1 - x_0} = \dfrac{\Delta y}{\Delta x}$.

ii). *Instantaneous Rate of Change*: If (x_0, y_0) is a point on the graph of $y = f(x)$, then the instantaneous rate of change of y with respect to x at x_0 is $f'(x_0)$.

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

The Number ***e*** as a limit

i). $\lim_{n \to +\infty} \left(1 + \frac{1}{n}\right)^n = e$

ii). $\lim_{n \to 0} \left(1 + \frac{n}{1}\right)^{\frac{1}{n}} = e$

Rolle's Theorem

If f is continuous on $[a,b]$ and differentiable on (a,b) such that $f(a) = f(b)$, then there is at least one number c in the open interval (a,b) such that $f'(c) = 0$.

Mean Value Theorem

If f is continuous on $[a,b]$ and differentiable on (a,b), then there is at least one number c in (a,b) such that $\dfrac{f(b) - f(a)}{b - a} = f'(c)$.

Extreme-Value Theorem

If f is continuous on a closed interval $[a,b]$, then $f(x)$ has both a maximum and minimum on $[a,b]$.

To find the maximum and minimum values of a function $y = f(x)$, locate

1. the points where $f'(x)$ is zero or where $f'(x)$ fails to exist.
2. the end points, if any, on the domain of $f(x)$.

Note: These are the only candidates for the value of x where $f(x)$ may have a maximum or a minimum.

Let f be differentiable for $a < x < b$ and continuous for a $a \le x \le b$,

1. If $f'(x) > 0$ for every x in (a,b), then f is increasing on $[a,b]$.
2. If $f'(x) < 0$ for every x in (a,b), then f is decreasing on $[a,b]$.

0.5 Errors

Measurement is the basis of scientific study. All measurements are, however, approximate values (not true values) within the limitation of a measuring device, measuring environment, process of measurement and human error. We seek to minimize uncertainty and hence error to the extent possible.

Further, there is an important aspect of reporting measurement. It should be consistent, systematic and revealing in the context of accuracy and precision. We must understand that an error in basic quantities propagate through mathematical formula leading to compounding of errors and misrepresentation of quantities.

Errors are broadly classified into two categories:

- Systematic error
- Random error

A systematic error impacts "accuracy" of the measurement. Accuracy means how close is the measurement with respect to "true" value. A "true" value of a quantity is a measurement, when errors on all accounts are minimized. We should distinguish "accuracy" of measurement with "precision" of measurement, which is related to the ability of an instrument to measure values with greater details (divisions).

The measurement of a weight on a scale with marking in kg is 79 kg, whereas measurement of the same weight on a different scale having further divisions in hectogram is 79.3 kg. The later weighing scale is more precise. The precision of measurement of an instrument, therefore, is a function of the ability of an instrument to read smaller divisions of a quantity.

Summary:

1. True value of a quantity is an “unknown”. We cannot know the true value of a quantity, even if we have measured it by chance, as we do not know the exact value of error in measurement. We can only approximate true value with greater accuracy and precision.
2. An accepted “true” measurement of a quantity is a measurement, when errors on all accounts are minimized.
3. “Accuracy” means how close the measurement is with respect to “true” measurement. It is associated with systematic error.
4. “Precision” of measurement is related to the ability of an instrument to measure values in greater detail. It is associated with random error.

Systematic error

A systematic error results due to faulty measurement practices. An error of this type is characterized by deviation in one direction from the true value. Itt means that the error is introduced, which is either less than or greater than the true value. Systematic error impacts the accuracy of measurement – not the precision of the measurement.

Systematic error results from:

1. faulty instrument
2. faulty measuring process and
3. personal bias

Clearly, this type of error cannot be minimized or reduced by repeated measurements. A faulty machine, for example, will not improve accuracy of measurement by repeating measurements.

Instrument error

A zero error, for example, is an instrument error, which is introduced in the measurement consistently in one direction. A zero error results when the zero mark of the scale does not match with pointer. We can realize this with the weighing instrument (scale) we use in our home. Often, the pointer is off the zero mark of the scale. Moreover, the scale may in itself be not uniformly marked or may not be properly calibrated. In vernier calipers, the nine divisions of main scale should be exactly equal to ten divisions of vernier scale. In a nutshell, we can say that the instrument error occurs due to faulty design of the instrument. We can minimize this error by replacing the instrument or by making a change in the design of the instrument.

Procedural error

A faulty measuring process may include an inappropriate physical environment, procedural mistakes and a lack of understanding of the process of measurement. For example, if we are studying the magnetic effect of current, then it would be erroneous to conduct the experiment in a place where strong currents are flowing nearby. Similarly, while taking the temperature of a human body, it is important to know which of the human parts is more representative of body temperature.

This type of error can be minimized by periodically assessing the measurement process and improvising the system. Consulting a subject expert or simply conducting an audit of the measuring process is also helpful in light of new facts and advancements.

Personal bias

A personal bias is introduced by human habits, which are not conducive for accurate measurement. Consider, for example, the reading habit of a person. He or she may have the habit of reading scales from an inappropriate distance and from an oblique direction. The measurement, therefore, includes error because of parallax.

Parallax

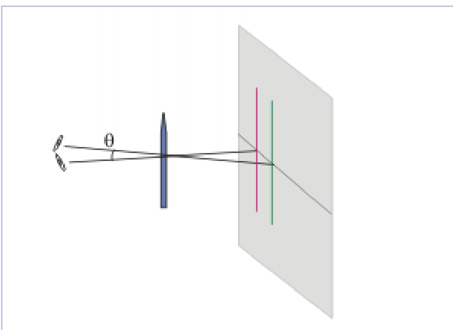


Figure 0.1: The position of pencil changes with respect to a mark on the background.

We can appreciate the importance of parallax by just holding a finger (pencil) in the hand, which is stretched horizontally. We keep the finger in front of our eyes against some reference marking in the background. Now, we look at the finger by closing one eye at a time and note the relative displacement of the finger with respect to the mark in the static background. We can do this experiment any time as shown in the figure above. The parallax results due to the angle at which we look at the object.

It is important that we read position of a pointer or a needle on a scale normally to avoid error because of parallax.

Commented [1]: Should this be "read the position"?

Parallax

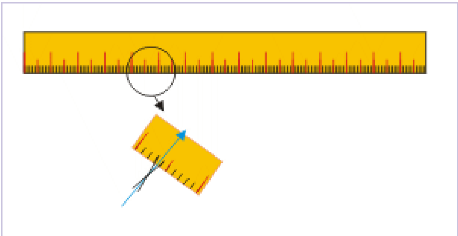


Figure 0.2: Parallax error is introduced as we may read values at an angle.

Random errors

Random error unlike systematic error is not unidirectional. Some of the measured values are greater than true value; some are less than true value. The errors introduced are sometimes positive and sometimes negative with respect to true value. It is possible to minimize this type of error by repeating measurements and applying statistical techniques to get closer value to the true value.

Another distinguishing aspect of random error is that it is not biased. It is there because of the limitation of the instrument in hand and the limitation on the part of human ability. No human being can repeat an action in exactly the same manner. Hence, it is likely that the same person reports different values with the same instrument, which measures the quantity correctly.

Least count error

Least count error results due to the inadequacy of resolution of the instrument. We can understand this in the context of least count of a measuring device. The least count of a device is equal to the smallest division on the scale. Consider the meter scale that we use. What is its least count? Its smallest division is in millimeter (mm). Hence, its least count is 1 mm i.e. 10^{-3} m i.e. 0.001 m. Clearly, this meter scale can be used to measure length from 10^{-3} m to 1 m. It is worth to know that least count of a vernier scale is 10^{-4} m and that of screw gauge and spherometer is 10^{-5} m.

Returning to the meter scale, we have the dilemma of limiting ourselves to the exact measurement up to the precision of marking or should be limited to a step before. For example, let us read the measurement of a piece of a given rod. One end of the rod exactly matches with the zero of scale. The other end lies at the smallest markings at 0.477 m (= 47.7 cm = 477 mm). We may argue that measurement should be limited to the marking, which can be definitely relied. If so, then we would report the length as 0.47 m, because we may not be definite about millimeter reading.

This is, however, unacceptable, as we are sure that length consists of some additional length – only thing that we may err as the reading might be 0.476 m or 0.478 m instead of 0.477 m. There is a definite chance of error due to limitation in reading such small divisions. We would be more precise and accurate by reporting measurement as 0.477 ± some agreed level of anticipated error. Generally, the accepted level of error in reading the smallest division is considered to be half of the least count. Hence, the reading would be:

$\Rightarrow x=0.477\pm 0.0005m$

If we report the measurement in centimeter,

$\Rightarrow x=47.7\pm 0.05cm$

If we report the measurement in millimeter,

$\Rightarrow x=477\pm 0.5mm$

Mean value of measurements

It has been pointed out that random error, including that of least count error, can be minimized by repeating measurements. It is so because errors are not unidirectional. If we take average of the measurements from the repeated measurements, it is likely that we minimize error by canceling out errors in opposite directions.

Here, we are implicitly assuming that measurement is free of “systematic errors”. The averaging of the repeated measurements, therefore, gives the best estimate of “true” value. As such, average or mean value (a_m) of the measurements (excluding "of- beat" measurements) is the notional “true” value of the quantity being measured. In fact, it is reported as true value, being our best estimate.

Error Propagation

In this module, we shall introduce some statistical analysis techniques to improve our understanding about error and enable reporting of error in the measurement of a quantity. There are three related approaches, which involves measurement of:

Absolute error

Relative error

Percentage error

Absolute error

The absolute error is the magnitude of error as determined from the difference of measured value from the mean value of the quantity. The important thing to note here is that absolute error is concerned with the magnitude of error – not the direction of error. For a particular nth measurement,

$$|\Delta x_n| = |x_n - x_n|$$

where "x_m" is the mean or average value of measurements and "x_n" is the nth instant of measurement.

In order to calculate a few absolute values, we consider a set of measured data for the length of a given rod. Note that we are reporting measurements in centimeter.

$x_1 = 47.7$ cm, $x_2 = 47.5$ cm, x_3 =47.8 cm, $x_4 = 47.4$ cm and $x_5 = 47.7$ cm

The mean value of length is,

$\Rightarrow x_m = 47.62$ cm

It is evident from the individual values that the least count of the scale (smallest division) is 0.001 m = 0.1 cm. For this reason, we limit mean value to the first decimal place. Hence, we round off the last but one digit as:

x_m=47.6cm

This is the mean or true value of the length of the rod. Now, absolute error of each of the five measurements are,

$|\Delta x_1|=|x_m - x_1| = |47.6-47.7| = |-0.1 = .1$ cm

$|\Delta x_2|=|x_m - x_2| = |47.6-47.5| = |0.1| =0.1$ cm

$|\Delta x_3|=|x_m - x_3| = |47.6-47.8| = |-0.2| = 0.2$ cm

$|\Delta x_4|=|x_m - x_4| = |47.6-47.4| = |0.2| = 0.2$ cm

$|\Delta x_5|=|x_m - x_5| = |47.6-47.7| = |-0.1| = 0.1$ cm

Mean absolute error

Earlier, it was stated that a quantity is measured with a range of error specified by half the least count. This is a generally accepted range of error. Here, we shall work to calculate the range of the error, based on the actual measurements and not go by any predefined range of error as that of generally accepted range of error. This means that we want to determine the range of error, which is based on the deviations in the reading from the mean value.

Absolute error associated with each measurement tells us how far the measurement can be off the mean value. The absolute errors calculated, however, may be different. Now the question is, which of the absolute errors should be taken for our consideration? We take the average of the absolute error:

The value of measurement, now, will be reported with the range of error as:

$x = x_m \pm \Delta x_m$

Extending this concept of defining range to the earlier example, we have,

$\Rightarrow \Delta x_m = 0.1$cm.

We should note here that we have rounded the result to reflect that the error value that has same precision as that of measured value. The value of the measurement with the range of error, then, is :

$\Rightarrow x = 47.6 \pm 0.1$cm

A plain reading of above expression is “the length of the rod lies in between 47.5 cm and 47.7 cm”. For all practical purpose, we shall use the value of x = 47.6 cm with caution, in that this quantity involves an error of the magnitude of “0.1 cm” in either direction.

Chapter 1

1.0 Objectives

At the end of this lesson, students should be able to,

1. Identify the scope of physics.
2. Calculate the order of magnitude of a quantity.
3. Use length, mass, and timescales in calculations.
4. Identify the relationships among models, theories, and laws.
5. Apply SI units.
6. Identify estimations.
7. Identify the relationship among the concepts of accuracy, precision, uncertainty, and discrepancy.
8. Calculate the percent uncertainty of a measurement.
9. Determine the uncertainty of the result of a computation involving quantities with given uncertainties.
10. Identify significant numbers in calculations.

1.1 Introduction

Physics involves the study of everything in physical existence from subatomic particles to galaxies. The study of physics shares common approaches and practices. The principles and methods used in physics can be applied in other sciences, too.

1.2 The Scope of Physics

The universe serves as an excellent starting point to understand the fundamental forces that hold the universe together and are useful in daily activities. Living beings, all matter around us, technology, professions, and other objects all rely on the principles of physics to function. The laws of physics connect seemingly unrelated topics. The study of matter and various forms of energy reveals how everything that surrounds us relates to one another. The scope of physics is concerned with describing the interactions of energy, matter, space, and time to uncover the fundamental mechanisms underlying every phenomenon. Physics emphasizes the use of a small number of quantitative laws that can be useful in other fields. Understanding the principles of physics can provide a greater understanding of the interconnectedness of everything we can see and know in this universe.

Understanding the basic laws of physics and their applications can help us make informed decisions in our daily lives. Moreover, physics plays a crucial role in many fields, ranging from engineering to medicine. It helps scientists and engineers design new technologies and structures, from bridges and buildings to advanced medical equipment. Physics also provides a fundamental framework for understanding natural phenomena, such as earthquakes, weather patterns, and the behavior of subatomic particles.

In summary, studying physics can provide valuable insights into the workings of the world around us and contribute to our problem-solving skills. It can also serve as a foundation for many other sciences and professions, making it a vital area of study.

1.3 Order of Magnitude

The order of magnitude is a way to estimate the scale of a number based on the power of 10 that most closely approximates it. Each power of 10 represents a different order of magnitude.

The following are two ways to find the order of magnitude:

1. To find the order of magnitude of a number, take the base -10 logarithm of the number and round it to the nearest integer, then the order of magnitude of the number is simply the resulting power of 10.

   e.g. The order of magnitude of 800 is 10^3 because $\log_{10} 800 \approx 2.903$, which rounds to 3. Similarly, the order of magnitude of 450 is 10^3 because $\log_{10} 450 \approx 2.653$, which rounds to 3 as well. Therefore, the numbers 800 and 450 are of the same order of magnitude: 10^3. Nevertheless, the order of magnitude of 250 is 10^2 because $\log_{10} 250 \approx 2.397$, which rounds to 2.

2. Another easy way to find the order of magnitude is to first write the number in scientific notation and then check if the first factor is greater or less than $\sqrt{10} = 10^{0.5} \approx 3$. The idea is that $\sqrt{10} = 10^{0.5}$ is halfway between $1=10^0$ and $10 =10^1$ on a log base -10 scale. Therefore, if the first factor is less than $\sqrt{10}$, then we round it down to 1 and the order of magnitude is simply whatever power of 10 is required to write the number in scientific notation. On the other hand, if the first factor is greater than $\sqrt{10}$, then we round it up to 10 and the order of magnitude is one power of 10 higher than the power needed to write the number in scientific notation.

   e.g. The number 800 can be written in scientific notation as 8×10^2. Because 8 is bigger than $\sqrt{10} \approx 3$, we say the order of magnitude of 800 is $10^{2+1}=10^3$. The number 450 can be written as 4.5×10^2, so its order of magnitude is also 10^3 because 4.5 is greater than 3. However, 250 written in scientific notation is 2.5×10^2 and 2.5 is less than 3, so its order of magnitude is 10^2.

1.4 Models, Theories, and Laws in Physics

A model is a representation of an idea, a process, an object, or a system that describes the phenomena that is difficult or impossible to observe, display, experience, or comprehend directly. A model could be confirmed through experimental results; nevertheless, it is limited to describing certain aspects of a physical system.

Bohr model is an example of a model that is used to describe the arrangements of electrons in an atom, since electrons cannot be observed directly.

A theory is an explanation of patterns in nature that can be tested for assertion and supported by scientific evidence and repeated verifications. Some theories use models to support explanation. Newton's theory of gravity is an example of a theory. Since Newton's gravity can be observed, a model is not necessary to explain this phenomenon.

Whilst a model can be used to describe only a certain aspect of a phenomenon, a theory must describe all aspects of a physical system.

A law in physics includes concise language to describe general patterns in nature. These are supported by scientific evidence and repeated experiments.

Laws and theories are similar in that they are both scientific statements that result from a tested hypothesis and are supported by scientific evidence. Nevertheless, law is usually reserved for a concise and very general statement that describes phenomena in nature. Newton's second law of motion and the law of the conservation of energy are two examples of laws.

When a theory is expressed in concise statements, they can be considered laws. The major difference between a law and a theory is that a theory is much more complex and dynamic. A law describes a single action whereas a theory explains an entire group of related phenomena.

When an experiment does not confirm the predictions, then that theory or law can be considered wrong. Laws can never be confirmed with complete certainty because it is not possible to conduct all possible experiments to confirm a law for any scenario. In physics, we assume that all scientific laws and theories are valid until they are proven or observed to be wrong. When valid experiments prove a law or a theory to be wrong, more experiments are carried out to modify the law or the theory, or they are discarded.

1.5 Dimensions of Physical Quantities - Mass (M), Length (L), and Time (T), etc.

Mass (M), length (L), and time (T) are fundamental physical quantities. Other fundamental units are electric current (I), thermodynamic temperature (Θ), amount of substance (N), and luminous intensity (J). Since these units are not derived from each other they are independent of each other.

The following table provides a comparison of length, mass, and time.

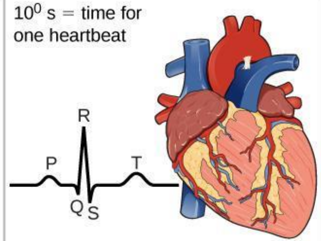
Length in Meters (m)	Masses in Kilograms (kg)	Time in Seconds (s)
10^{-15} m = diameter of proton	10^{-30} kg = mass of electron	10^{-22} s = mean lifetime of very unstable nucleus
10^{-14} m = diameter of large nucleus	10^{-27} kg = mass of proton	10^{-17} s = time for single floating-point operation in a supercomputer
10^{-10} m = diameter of hydrogen atom	10^{-15} kg = mass of bacterium	10^{-15} s = time for one oscillation of visible light
10^{-7} m = diameter of typical virus	10^{-5} kg = mass of mosquito	10^{-13} s = time for one vibration of an atom in a solid
10^{-2} m = pinky fingernail width	10^{-2} kg = mass of hummingbird	10^{-3} s = duration of a nerve impulse
10^{0} m = height of 4 year old child 	10^{0} kg = mass of liter of water 	10^{0} s = time for one heartbeat 
10^{2} m = length of football field	10^{2} kg = mass of person	10^{5} s = one day
10^{7} m = diameter of Earth	10^{19} kg = mass of atmosphere	10^{7} s = one year
10^{13} m = diameter of solar system	10^{22} kg = mass of Moon	10^{9} s = human lifetime
10^{16} m = distance light travels in a year (one light-year)	10^{25} kg = mass of Earth	10^{11} s = recorded human history
10^{21} m = Milky Way diameter	10^{30} kg = mass of Sun	10^{17} s = age of Earth
10^{26} m = distance to edge of observable universe	10^{53} kg = upper limit on mass of known universe	10^{18} s = age of the universe

Table 1.1: This table shows the orders of magnitude of length, mass, and time.

The dimension of any physical quantity expresses its dependence on the base quantities as a product of symbols (or powers of symbols) representing the base quantities. For example, a measurement of length is said to have dimension L or L^1, a measurement of mass has dimension M or M^1, and a measurement of time has dimension T or T^1. Like units, dimensions obey the rules of algebra. Thus, area is the product of two lengths and so has dimension L^2, or length squared. Similarly, volume is the product of three lengths and has dimension L^3, or length cubed. Speed has dimension length over time, L/T or LT^{-1}. Volumetric mass density has dimension M/L^3 or ML^{-3}, or mass over length cubed. In general, the dimension of any physical quantity can be written as $L^a M^b T^c I^d \Theta^e N^f J^g$ for some powers a, b, c, d, e, f, and g. We can write the

Commented [2]: Image source?

dimensions of a length in this form with a=1 and the remaining six powers all set equal to zero: $L^1 = L^1 M^0 T^0 I^0 \Theta^0 N^0 J^0$. Any quantity with a dimension that can be written so that all seven powers are zero (that is, its dimension is $L^0 M^0 T^0 I^0 \Theta^0 N^0 J^0$) is called dimensionless (or sometimes "of dimension 1," because anything raised to the zero power is one). Physicists often call dimensionless quantities pure numbers.

Physicists often use square brackets around the symbol for a physical quantity to represent the dimensions of that quantity. For example, if r is the radius of a cylinder and h is its height, then we write $[r]$=L and $[h]$=L to indicate the dimensions of the radius and height are both those of length, or L. Similarly, if we use the symbol A for the surface area of a cylinder and V for its volume, then $[A] = L^2$ and $[V] = L^3$. If we use the symbol m for the mass of the cylinder and ρ for the density of the material from which the cylinder is made, then $[m]$=M and $[\rho]=ML^{-3}$.

The importance of the concept of dimension arises from the fact that any mathematical equation relating physical quantities must be dimensionally consistent, which means the equation must obey the following rules:

- Every term in an expression must have the same dimensions; it does not make sense to add or subtract quantities of differing dimensions (think of the old saying: "You can't add apples and oranges"). In particular, the expressions on each side of the equality in an equation must have the same dimensions.
- The arguments of any of the standard mathematical functions such as trigonometric functions (such as sine and cosine), logarithms, or exponential functions that appear in the equation must be dimensionless. These functions require pure numbers as inputs and give pure numbers as outputs.

If either of these rules is violated, an equation is not dimensionally consistent and cannot possibly be a correct statement of physical law. This simple fact can be used to check for typos or algebra mistakes, to help remember the various laws of physics, and even to suggest the form that new laws of physics might take.

1.6 Units

In any system of units, the units for some physical quantities must be defined through a measurement process. These are called the base quantities for that system and their units are the system's base unit**s**. All other physical quantities can then be expressed as algebraic combinations of the base quantities. Each of these physical quantities is then known as a derived quantity and each unit is called a derived unit. The choice of base quantities is somewhat arbitrary, as long as they are independent of each other, and all other quantities can be derived from them. Typically, the goal is to choose physical quantities that can be measured accurately to a high precision as the base quantities. The reason for this is simple. Since the derived units can be expressed as algebraic combinations of the base units, they can only be as accurate and precise as the base units from which they are derived.

The International Standards Organization recommends using seven base quantities, which form the International System of Quantities (ISQ). These are the base quantities used to define the SI base units.

Most of the mechanical quantities can be expressed by using M, L, and T. The following table provides a list of the seven base quantity and base units:

Base Quantity	**Base Unit**
Length (L)	meter (m)
Mass (M)	kilogram (kg)
Time (T)	second (s)
Electrical current (I)	ampere (A)
Thermodynamic temperature (Θ)	kelvin (K)
Amount of substance (N)	mole (mol)
Luminous intensity (J)	candela (cd)

Table 1.2: Base quantity and base units

A physical quantity is defined either by specifying how it is measured or by stating how it is calculated from other measurements. As an example, distance and time are defined by specifying methods for measuring them, such as using a meter stick and a stopwatch. The average speed is defined by stating that it is calculated as the total distance traveled divided by the time of travel.

Measurements of physical quantities are expressed in terms of units, which are standardized values. For example, the length of a race, which is a physical quantity, can be expressed in units of meters or kilometers. Without standardized units, it would be extremely difficult for scientists to express and compare measured values in a meaningful way.

All derived quantities are formed from the base quantities given in Table 1.2. As an example, the geometric concept of area is always calculated as the product of two lengths. Thus, area is a derived quantity that can be expressed in terms of SI base units using square meters (m X m = m2). Similarly, volume is a derived quantity that can be expressed in cubic meters (m^3). Speed is length per time; so, in terms of SI base units, one could measure it in meters per second (m/s). Volume mass density (or just density) is mass per volume, which is expressed in terms of SI base units such as kilograms per cubic meter (kg/m^3). Angles can also be thought of as derived quantities because they can be defined as the ratio of the arc length subtended by two radii of a circle to the radius of the circle. This is how the radian is defined.

The following table provides some examples of derived units:

Quantity	**Symbol**	**Derived SI unit**	**Unit Abbreviation**	**Derivation Description**	**Derivation in base SI Units**
Area	A	square meter	m^2	length X width	m X m = m^2
Volume	V (noted that V is also used for voltage)	cubic meter	m^3	length X width X height	m X m X m = m^3
Density	D	kilograms/cubic meter	kg/m3	mass / volume	kg/m X m X m = kg m^{-3}
Concentration	c	moles/liter	mol/l	moles/volume	mol/l = mol l^{-1}
Velocity	v	meters/second	m/s	distance (length)/time	m / s = m s^{-1}
Acceleration	a	meters/second/second	m/s2	speed/time	ms^{-1}/ s = m.s^{-2}
Force	F	newton	N	mass X acceleration	kg X m/s^2 = kg ms^{-2}
Energy/Work /Heat	E	joule	J	force X distance (length)	Nm = kg ms^{-2} X m = kg m^2s^{-2}
Power/Radiant Flux	W	watt	J/s	energy/time	kg m^2 s^{-3}

Electric Charge/Quantity of Electricity	C	coulomb	A.s	current X time	A s

Table 1.3: Derived Quantity and Derived Units

Two major systems of units are used in the world: SI units (in French Système International d'Unités), also known as the metric system, and English units (also known as the customary or imperial system). English units were historically used in nations once ruled by the British Empire and are still widely used in the United States. English units may also be referred to as the foot–pound–second (fps) system, as opposed to the centimeter–gram–second (cgs) system. There is also another system of unites known as SAE units, named after the Society of Automotive Engineers. Products such as fasteners and automotive tools (for example, wrenches) that are measured in inches rather than metric units are referred to as SAE fasteners or SAE wrenches.

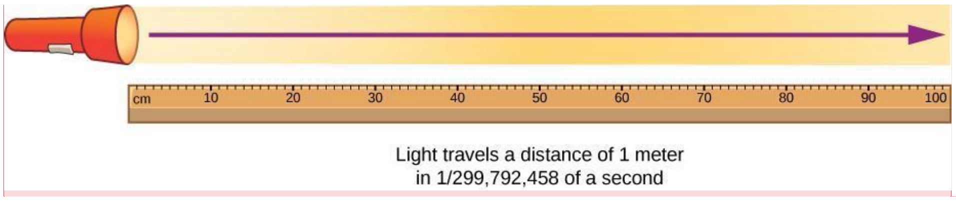
Except the United States, almost all other countries in the world now uses SI units as the standard. The metric system is also the standard system agreed on by scientists and mathematicians.

1.6.1 The second

The SI unit for time, the second (abbreviated s), has a long history. For many years it was defined as 1/86,400 of a mean solar day. More recently, a new standard was adopted to gain greater accuracy and to define the second in terms of a nonvarying or constant physical phenomenon (because the solar day is getting longer as a result of the very gradual slowing of Earth's rotation). Cesium atoms can be made to vibrate in a very steady way, and these vibrations can be readily observed and counted. In 1967, the second was redefined as the time required for 9,192,631,770 of these vibrations to occur.

1.6.2 The meter

The SI unit for length is the meter (abbreviated m); its definition has also changed over time to become more precise. The meter was first defined in 1791 as 1/10,000,000 of the distance from the equator to the North Pole. This measurement was improved in 1889 by redefining the meter to be the distance between two engraved lines on a platinum–iridium bar now kept near Paris. By 1960, it had become possible to define the meter even more accurately in terms of the wavelength of light, so it was again redefined as 1,650,763.73 wavelengths of orange light emitted by krypton atoms. In 1983, the meter was given its current definition (in part for greater accuracy) as the distance light travels in a vacuum in 1/299,792,458 of a second. This change came after knowing the speed of light to be exactly 299,792,458 m/s.



Commented [3]: Source?

Figure 1.1 The meter is defined to be the distance light travels in 1/299,792,458 of a second in a vacuum. Distance traveled is speed multiplied by time.

1.6.3 The kilogram

The SI unit for mass is the kilogram (abbreviated kg); From 1795–2018 it was defined to be the mass of a platinum–iridium cylinder kept with the old meter standard at the International Bureau of Weights and Measures near Paris. However, this cylinder has lost roughly 50 micrograms since it was created. Because this is the standard, this has shifted how we defined a kilogram. Therefore, a new definition was adopted in May 2019 based on the Planck constant and other constants which will never change in value. The kilogram is measured on a Kibble balance. When a weight is placed on a Kibble balance, an electrical current is produced that is proportional to Planck's constant. Since Planck's constant is defined, the exact current measurements in the balance define the kilogram.

1.7 Metric Prefixes

SI units are part of the metric system, which is convenient for scientific and engineering calculations because the units are categorized by factors of 10. The table 1.4 lists the metric prefixes and symbols used to denote various factors of 10 in SI units. For example, a centimeter is one-hundredth of a meter (in symbols, 1 cm = 10^{-2} m) and a kilometer is a thousand meters (1 km = 10^{3} m). Similarly, a megagram is a million grams (1 Mg = 10^{6} g), a nanosecond is a billionth of a second (1 ns = 10^{-9} s), and a terameter is a trillion meters (1 Tm = 10^{12} m).

Prefix	Symbol	Meaning	Prefix	Symbol	Meaning
yotta-	Y	10^{24}	yocto-	y	10^{-24}
zetta-	Z	10^{21}	zepto-	z	10^{-21}
exa-	E	10^{18}	atto-	a	10^{-18}
peta-	P	10^{15}	femto-	f	10^{-15}
tera-	T	10^{12}	pico-	p	10^{-12}
giga-	G	10^{9}	nano-	n	10^{-9}
mega-	M	10^{6}	micro-	μ	10^{-6}
kilo-	k	10^{3}	milli-	m	10^{-3}
hecto-	h	10^{2}	centi-	c	10^{-2}
deka-	da	10^{1}	deci-	d	10^{-1}

Table 1.4: Metric Prefixes for Powers of 10 and Their Symbols

The only rule when using metric prefixes is that you cannot “double them up.” For example, if you have measurements in petameters (1 Pm = 10^{15} m), it is not proper to talk about megagigameters, although $10^6 \times 10^9 = 10^{15}$. In practice, the only time this becomes a bit confusing is when discussing masses. As we have seen, the base SI unit of mass is the kilogram (kg), but metric prefixes need to be applied to the gram (g), because we are not allowed to “double-up” prefixes. Thus, a thousand kilograms (10^3 kg) is written as a megagram (1 Mg) since

$$10^3 \text{kg} = 10^3 \times 10^3 \text{g} = 10^6 \text{g} = 1 \text{Mg}$$

Incidentally, 10^3 kg is also called a metric ton, abbreviated t. This is one of the units outside the SI system considered acceptable for use with SI units.

The metric systems have the advantage that conversions of units involve only powers of 10. There are 100 cm in 1 m, 1000 m in 1 km, and so on. In nonmetric systems, such as the English system of units, the relationships are not as simple - there are 12 in. in 1 ft, 5280 ft in 1 mi, and so on.

Another advantage of metric systems is that the same unit can be used over extremely large ranges of values simply by scaling it with an appropriate metric prefix. The prefix is chosen by the order of magnitude of physical quantities commonly found in the task at hand. For example, distances in meters are suitable in construction, whereas distances in kilometers are appropriate for air travel, and nanometers are convenient in optical design. With the metric system there is no need to invent new units for applications. Instead, we rescale the units with which we are already familiar.

An example on using metric prefixes:

Restate the mass 1.93×10^{13}kg using a metric prefix such that the resulting numerical value is bigger than one but less than 1000.

Strategy

Since we are not allowed to “double-up” prefixes, we first need to restate the mass in grams by replacing the prefix symbol k with a factor of 10^3. Then, we should see which two prefixes are closest to the resulting power of 10 when the number is written in scientific notation. We use whichever of these two prefixes gives us a number between one and 1000.

Solution

Replacing the k in kilogram with a factor of 10^3, we find that

$$1.93 \times 10^{13} \text{kg} = 1.93 \times 10^{13} \times 10^3 \text{g} = 1.93 \times 10^{16} \text{g}$$

10^{16} is between “peta-” (10^{15}) and “exa-” (10^{18}). If we use the “peta-” prefix, then we find that 1.93×10^{16}g = 1.93×10^1Pg, since 16 =1+15. Alternatively, if we use the “exa-” prefix we find that

1.93×10^{16}g $=1.93\times10^{-2}$ Eg, since $16 = -2+18$. Because the problem asks for the numerical value between one and 1000, we use the "peta−" prefix and the answer is 19.3 Pg.

1.8 Conversion of Units

Suppose we want to convert 80 m to kilometers. The first thing to do is to list the units you have and the units to which you want to convert. In this case, we have units in meters, and we want to convert to kilometers. Next, we need to determine a conversion factor relating to meters to kilometers. A conversion factor is a ratio that expresses how many times of a particular unit are equal to another unit. For example, there are 12 in. in 1 ft, 1609 m in 1 mi, 100 cm in 1 m, 60 s in 1 min, and so on. In this case, we know that there are 1000 m in 1 km. Now we can set up our unit conversion. We write the units we have and then multiply them by the conversion factor, so the units cancel out, as shown:

$$80\text{m} \times (1\text{km}/1000\text{m}) = 0.080\text{km}$$

Note that the unwanted meter unit cancels, leaving only the desired kilometer unit. You can use this method to convert between any type of unit. Now, the conversion of 80 m to kilometers is simply the use of a metric prefix, so we can get the same answer just as easily by noting that

$$80\text{m} = 8.0\times10^{1}\text{ m} = 8.0\times10^{-2}\text{ km} = 0.080\text{km}$$

since "kilo-" means 10^3 and $1 = -2 + 3$. However, using conversion factors is handy when converting between units that are not metric or when converting between derived units, as the following examples illustrate.

An example of converting nonmetric units to metric is given below:

The distance from the university to home is 10 mi and it usually takes 20 min to drive this distance. Calculate the average speed in meters per second (m/s). (*Note:* Average speed is distance traveled divided by time of travel.)

Strategy

Firstly, we calculate the average speed using the given units, then we can get the average speed into the desired units by picking the correct conversion factors and multiplying them. The correct conversion factors are those that cancel the unwanted units and leave the desired units in their place. In this case, we want to convert miles to meters, so we need to know the fact that there are 1609 m in 1 mi. We also want to convert minutes to seconds, so we use the conversion of 60 s in 1 min.

Solution

1. Calculate average speed. Average speed is distance traveled divided by time of travel. (Take this definition as a given for now. Average speed and other motion concepts are covered in later chapters.) In equation form,

$$\text{Average speed} = \text{Distance} / \text{Time}.$$

2. Substitute the given values for distance and time:

$$\text{Average speed} = 10 \text{ mi} / 20 \text{ min} = 0.50 \text{ mi} / \text{min}.$$

3. Convert miles per minute to meters per second by multiplying by the conversion factor that cancels miles and leave meters, and also by the conversion factor that cancels minutes and leave seconds:

$$0.50 \text{ mile} / \text{min} \times (1609 \text{ m} / 1 \text{mile}) \times (1 \text{min} / 60 \text{ s}) = (0.50)\ (1609)\ 60 \text{ m/s} = 13 \text{ m/s}.$$

Important Information

Check the answer in the following ways:

1. Be sure the units in the unit conversion cancel correctly. If the unit conversion factor was written upside down, the units do not cancel correctly in the equation. We see the “miles” in the numerator in 0.50 mi/min cancels the “mile” in the denominator in the first conversion factor. Also, the “min” in the denominator in 0.50 mi/min cancels the “min” in the numerator in the second conversion factor.
2. Check that the units of the final answer are the desired units. The problem asked us to solve for average speed in units of meters per second and, after the cancellations, the only units left are a meter (m) in the numerator and a second (s) in the denominator, so we have indeed obtained these units.

1.9 Estimation

On many occasions, physicists, other scientists, and engineers need to make estimates for a particular quantity. Other terms sometimes used are guesstimates, order-of-magnitude approximations, back-of-the-envelope calculations, or Fermi calculations. (The physicist Enrico Fermi mentioned earlier was famous for his ability to estimate various kinds of data with surprising precision.)

Will that piece of equipment fit in the back of the car or do we need to rent a truck? How long will this download take? About how large a current will there be in this circuit when it is turned on? How many houses could a proposed power plant actually power if it is built?

Note that estimating does not mean guessing a number or a formula at random. Rather, estimation means using prior experience and sound physical reasoning to arrive at a rough idea of a quantity's value. Because the process of determining a reliable approximation usually involves the identification of correct physical principles and a good guess about the relevant variables, estimating is very useful in developing physical intuition. Estimates also allow us to perform "sanity checks" on calculations or policy proposals by helping us rule out certain scenarios or unrealistic numbers. They allow us to challenge others (as well as ourselves) in our efforts to learn truths about the world.

Many estimates are based on formulas in which the input quantities are known only to a limited precision. As you develop physics problem-solving skills (which are applicable to a wide variety of fields), you also will develop skills at estimating. You develop these skills by thinking quantitatively and being willing to take risks. As with any skill, experience helps. Familiarity with dimensions and units, and the scales of base quantities also helps.

To make some progress in estimating, you need to have some definite ideas about how variables may be related. The following strategies may help you in practicing the art of estimation:

- **Get big lengths from smaller lengths.** When estimating lengths, remember that anything can be a ruler. Thus, imagine breaking a big thing into smaller things, estimate the length of one of the smaller things, and multiply to get the length of the big thing. For example, to estimate the height of a building, first count how many floors it has. Then, estimate how big a single floor is by imagining how many people would have to stand on each other's shoulders to reach the ceiling. Last, estimate the height of a person. The product of these three estimates is your estimate of the height of the building. It helps to have memorized a few length scales relevant to the sorts of problems you find yourself solving. For example, knowing some of the length scales in might come in handy. Sometimes it also helps to do this in reverse—that is, to estimate the length of a small thing, imagine a bunch of them making up a bigger thing. For example, to estimate the thickness of a sheet of paper, estimate the thickness of a stack of paper and then divide by the number of pages in the stack. These same strategies of breaking big things into smaller things or aggregating smaller things into a bigger thing can sometimes be used to estimate other physical quantities, such as masses and times.
- **Get areas and volumes from lengths**. When dealing with an area or a volume of a complex object, introduce a simple model of the object such as a sphere or a box. Then, estimate the linear dimensions (such as the radius of the sphere or the length, width, and height of the box) first, and use your estimates to obtain the volume or area from standard geometric formulas. If you happen to have an estimate of an object's area or volume, you can also do the reverse; that is, use standard geometric formulas to get an estimate of its linear dimensions.
- **Get masses from volumes and densities**. When estimating masses of objects, it can help first to estimate its volume and then to estimate its mass from a rough estimate of its average density (recall, density has dimension mass over length cubed, so mass is density times volume). For this, it helps to remember that the density of air is around 1 kg/m^3, the density of water is 10^3 kg/m^3, and the densest everyday solids max out at around 10^4 kg/m^3. Asking yourself whether an object floats or sinks in either air or water gets you a ballpark estimate of its density. You can also do this the other way around; if you have an

estimate of an object's mass and its density, you can use them to get an estimate of its volume.

- **If all else fails, bound it.** For physical quantities for which you do not have a lot of intuition, sometimes the best you can do is think something like: Well, it must be bigger than this and smaller than that. For example, suppose you need to estimate the mass of a moose. Maybe you have a lot of experience with moose and know their average mass offhand. If so, great. But for most people, the best they can do is to think something like: It must be bigger than a person (of order 10^2 kg) and less than a car (of order 10^3 kg). If you need a single number for a subsequent calculation, you can take the geometric mean of the upper and lower bound—that is, you multiply them together and then take the square root. For the moose mass example, this would be

$$(10^2 \times 10^3)^{0.5} = 10^{2.5} = 10^{0.5} \times 10^2 \approx 3 \times 10^2 \text{kg}.$$

  The tighter the bounds, the better. Also, no rules are unbreakable when it comes to estimation. If you think the value of the quantity is likely to be closer to the upper bound than the lower bound, then you may want to bump up your estimate from the geometric mean by an order or two of magnitude.

- **One "sig. fig." is fine (significant figures are discussed later in this chapter)**. There is no need to go beyond one significant figure, or one digit in the coefficient of an expression in scientific notation, when doing calculations to obtain an estimate. In most cases, the order of magnitude is good enough. The goal is just to get in the ballpark figure, so keep the arithmetic as simple as possible.
- **Ask yourself: Does this make any sense?** Last, check to see whether your answer is reasonable. How does it compare with the values of other quantities with the same dimensions that you already know or can look up easily? If you get some wacky answer (for example, if you estimate the mass of the Atlantic Ocean to be bigger than the mass of Earth, or some time span to be longer than the age of the universe), first check to see whether your units are correct. Then, check for arithmetic errors. Then, rethink the logic you used to arrive at your answer. If everything checks out, you may have just proved that some slick new idea is actually bogus.

e.g. : Estimate the total mass of the oceans on Earth.

Strategy

We know the density of water is about 10^3 kg/m^3, so we start with the advice to "get masses from densities and volumes." Thus, we need to estimate the volume of the planet's oceans. Using the advice to "get areas and volumes from lengths," we can estimate the volume of the oceans as surface area times average depth, or $V = AD$. Assume that we know the diameter of Earth. Also, we know that most of Earth's surface is covered in water, so we can estimate the surface area of the oceans as being roughly equal to the surface area of the planet. By following the advice to "get areas and volumes from lengths" again, we can approximate Earth as a sphere and use the formula for the surface area of a sphere of diameter d - that is, $A=\pi d2,$ to estimate the surface area of the oceans. Now we just need to estimate the average depth of the oceans. For this, we use the advice: "If all else fails, bound it." We happen to know the deepest points in the ocean are around 10 km and that it is not uncommon for the ocean to be deeper than 1 km, so we take the average depth to be around $(10^3 \times 10^4)^{0.5} \approx 3 \times 10^3$m. Now we just need to put it all together, heeding the advice that "one 'sig. fig.' is fine."

Solution

We estimate the surface area of Earth (and hence the surface area of Earth's oceans) to be roughly,

$$A = \pi d^2 = \pi\,(10^7\text{m})^2 \approx 3 \times 10^{14}\text{m}^2.$$

Next, using our average depth estimate of $D = 3 \times 10^3$m, which was obtained by bounding, we estimate the volume of Earth's oceans to be

$$V = AD = (3 \times 10^{14}\text{m}^2)(3 \times 10^3\text{m}) = 9 \times 10^{17}\text{m}^3.$$

Last, we estimate the mass of the world's oceans to be

$$M = \rho V = (10^3\text{kg/m}^3)\,(9 \times 10^{17}\text{m}^3) = 9 \times 10^{20}\text{kg}.$$

Thus, we estimate that the order of magnitude of the mass of the planet's oceans is 10^{21} kg.

1.10 Accuracy and Precision

Science is based on observation and experiment—that is, on measurements. **Accuracy** is how close a measurement is to the accepted reference value for that measurement. For example, let's say we want to measure the length of standard printer paper. The packaging in which we purchased the paper states that it is 11.0 in. long. We then measure the length of the paper three times and obtain the following measurements: 11.1 in., 11.2 in., and 10.9 in. These measurements are quite accurate because they are very close to the reference value of 11.0 in. In contrast, if we had obtained a measurement of 12 in., our measurement would not be very accurate. Notice that the concept of accuracy requires that an accepted reference value be given.

The precision of measurements refers to how close the agreement is between repeated independent measurements (which are repeated under the same conditions). Consider the example of the paper measurements. The precision of the measurements refers to the spread of the measured values. One way to analyze the precision of the measurements is to determine the range, or difference, between the lowest and the highest measured values. In this case, the lowest value was 10.9 in. and the highest value was 11.2 in. Thus, the measured values deviated from each other by, at most, 0.3 in. These measurements were relatively precise because they did not vary too much in value. However, if the measured values had been 10.9 in., 11.1 in., and 11.9 in., then the measurements would not be very precise because there would be significant variation from one measurement to another. Notice that the concept of precision depends only on the actual measurements acquired and does not depend on an accepted reference value.

The measurements in the paper example are both accurate and precise, but in some cases, measurements are accurate but not precise, or they are precise but not accurate. Let's consider an example of a GPS attempting to locate the position of a restaurant in a city. Think of the restaurant location as existing at the center of a bull's-eye target and think of each GPS attempt to locate the restaurant as a black dot. In Figure 1.2 (a), we see the GPS measurements are spread out far apart from each other, but they are all relatively close to the actual location of the restaurant at the center of the target. This indicates a low-precision, high-accuracy measuring system. However, in Figure (b), the GPS measurements are concentrated quite closely to one another, but they are far away from the target location. This indicates a high-precision, low-accuracy measuring system.

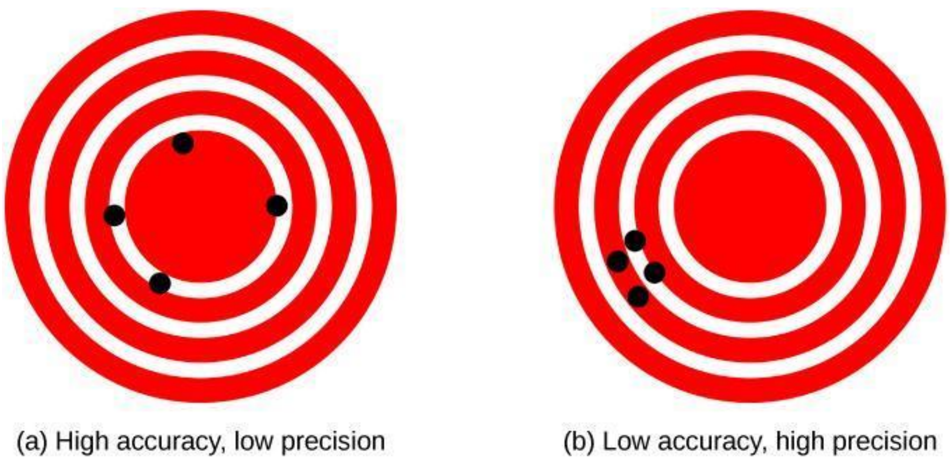


Figure 1.2: A GPS attempts to locate a restaurant at the center of the bull's-eye. The black dots represent each attempt to pinpoint the location of the restaurant. (a) The dots are spread out quite far apart from one another, indicating low precision, but they are each rather close to the actual location of the restaurant, indicating high accuracy. (b) The dots are concentrated rather closely to one another, indicating high precision, but they are rather far away from the actual location of the restaurant, indicating low accuracy. (credit a and credit b: modification of works by "DarkEvil"/Wikimedia Commons)

Accuracy, Precision, Uncertainty, and Discrepancy

The precision of a measuring system is related to the uncertainty in the measurements whereas the accuracy is related to the discrepancy from the accepted reference value. Uncertainty is a quantitative measure of how much your measured values deviate from one another. There are many different methods of calculating uncertainty, each of which is appropriate to different situations. Some examples include taking the range (that is, the largest minus the smallest) or finding the standard deviation of the measurements. Discrepancy (or "measurement error") is the difference between the measured value and a given standard or expected value. If the measurements are not very precise, then the uncertainty of the values is high. If the measurements are not very accurate, then the discrepancy of the values is high.

Recall our example of measuring paper length; we obtained measurements of 11.1 in., 11.2 in., and 10.9 in., and the accepted value was 11.0 in. We might average the three measurements to say our best guess is 11.1 in.; in this case, our discrepancy is 11.1 – 11.0 = 0.1 in., which provides a quantitative measure of accuracy. We might calculate the uncertainty in our best guess by using half of the range of our measured values: 0.15 in. Then we would say the length of the paper is 11.1 in. plus or minus 0.15 in. The uncertainty in a measurement, A, is often denoted as δA (read "delta A"), so the measurement result would be recorded as $A \pm \delta A$. Returning to our paper example, the measured length of the paper could be expressed as 11.1 ± 0.15 in. Since the discrepancy of 0.1 in. is less than the uncertainty of 0.15 in., we might say the measured value agrees with the accepted reference value to within experimental uncertainty.

Some factors that contribute to uncertainty in a measurement include the following:

- Limitations of the measuring device
- The skill of the person taking the measurement
- Irregularities in the object being measured
- Any other factors that affect the outcome (highly dependent on the situation)

In our example, such factors contributing to the uncertainty could be the smallest division on the ruler is 1/16 in., the person using the ruler has bad eyesight, the ruler is worn down on one end, or one side of the paper is slightly longer than the other. At any rate, the uncertainty in a measurement must be calculated to quantify its precision. If a reference value is known, it makes sense to calculate the discrepancy as well to quantify its accuracy.

1.11 Percent of uncertainty or percent of error

Another method of expressing uncertainty is as a percent of the measured value. If a measurement A is expressed with uncertainty δA, the percent uncertainty is defined as

Percent uncertainty = $(\delta A/A) \times 100\%$.

e.g. : **Calculating Percent Uncertainty: A Bag of Apples**

A grocery store sells 5-lb bags of apples. Let's say we purchase four bags during the course of a month and weigh the bags each time. We obtain the following measurements:

- Week 1 weight: 4.8 lb
- Week 2 weight: 5.3 lb
- Week 3 weight: 4.9 lb
- Week 4 weight: 5.4 lb

We then determine the average weight of the 5-lb bag of apples is 5.1 ± 0.3 lb from using half of the range. What is the percent uncertainty of the bag's weight?

Strategy

First, observe that the average value of the bag's weight, *A*, is 5.1 lb. The uncertainty in this value, δA, is 0.3 lb. We can use the following equation to determine the percent uncertainty of the weight:

Percent uncertainty = (δA/A) ×100%.

Solution

Substitute the values into the equation:

Percent uncertainty = (δA/A) × 100% = (0.3lb/5.1lb) × 100% = 5.9% ≈ 6%.

Therefore, we can conclude the average weight of a bag of apples from this store is 5.1 lb ± 6%. Notice the percent uncertainty is dimensionless because the units of weight in δA=0.2 lb canceled those in *A* = 5.1 lb when we took the ratio.

1.12 Significant Figures

An important factor in the precision of measurements involves the precision of the measuring tool. In general, a precise measuring tool is one that can measure values in very small increments. For example, a standard ruler can measure length to the nearest millimeter whereas a caliper can measure length to the nearest 0.01 mm. The caliper is a more precise measuring tool because it can measure extremely small differences in length. The more precise the measuring tool, the more precise the measurements.

When we express measured values, we can only list as many digits as we measured initially with our measuring tool. For example, if we use a standard ruler to measure the length of a stick, we may measure it to be 36.7 cm. We can't express this value as 36.71 cm because our measuring tool is not precise enough to measure a hundredth of a centimeter. It should be noted that the last digit in a measured value has been estimated in some way by the person performing the measurement. For example, the person measuring the length of a stick with a ruler notices the stick length seems to be somewhere in between 36.6 cm and 36.7 cm, and he or she must estimate the value of the last digit. Using the method of significant figures, the rule is that the last digit written down in a measurement is the first digit with some uncertainty. To determine the number of significant digits in a value, start with the first measured value at the left and count the number of digits through the last digit written on the right. For example, the measured value 36.7 cm has three digits, or three significant figures. Significant figures indicate the precision of the measuring tool used to measure a value.

All numbers from 1 – 9 are considered significant in a value. When it comes to zeros, we follow the rules given below:

Zeros

Special consideration is given to zeros when counting significant figures.

1. The leading zeros at the beginning of a value are not considered significant with or without decimal in the value.
   The zeros in 0.053 are not significant because they are placeholders that locate the decimal point. There are two significant figures in 0.053.
2. The trapped zeros in the middle of a value are considered significant.
   The zeros in 10.053 are not placeholders; they are significant. This number has five significant figures.
3. The trailing zeros at the end of a number are significant only if there is a decimal in the value. The zeros in 1300 may or may not be significant, depending on the style of writing numbers. They could mean the number is known to the last digit or they could be placeholders. So, 1300 could have two, three, or four significant figures. To avoid this ambiguity, we should write 1300 in scientific notation as 1.3×10^3, 1.30×10^3, or 1.300×10^3, depending on whether it has two, three, or four significant figures. Similarly, 1300. has four significant figures since there is a decimal at the end, and 1300.00 has six significant figures.

1.121 Significant figures in calculations

When combining measurements with different degrees of precision with the mathematical operations of addition, subtraction, multiplication, or division, then **the number of significant digits in the final answer can be no greater than the number of significant digits in the least-precise measured value.** There are two different rules, one for multiplication and division and the other for addition and subtraction.

1. *For multiplication and division, the result should have the same number of significant figures as the quantity with the least number of significant figures entering into the calculation*. For example, the area of a circle can be calculated from its radius using $A = \pi r^2$. Let's see how many significant figures the area has if the radius has only two—say, $r = 1.2$ m. Using a calculator with an eight-digit output, we would calculate

   $A = \pi r^2 = (3.1415927\ldots) \times (1.2\text{m})^2 = 4.5238934\text{m}^2$.

   But because the radius has only two significant figures, it limits the calculated quantity to two significant figures, or

   $A = 4.5\text{m}^2$,

   although π is good to at least eight digits.

2. **For addition and subtraction, the answer can contain no more decimal places than the least-precise measurement.** Suppose we buy 7.56 kg of potatoes in a grocery store as measured with a scale with precision 0.01 kg, then we drop off 6.052 kg of potatoes at your laboratory as measured by a scale with precision 0.001 kg. Then, we go home and add 13.7 kg of potatoes as measured by a bathroom scale with precision 0.1 kg. How

many kilograms of potatoes do we now have and how many significant figures are appropriate in the answer? The mass is found by simple addition and subtraction:

7.56kg – 6.052kg + 13.7kg = 15.208kg

Next, we identify the least-precise measurement: 13.7 kg. This measurement is expressed to the 0.1 decimal place, so our final answer must also be expressed to the 0.1 decimal place. Thus, the answer is rounded to the tenths place, giving us 15.2 kg.

Significant figures in this text

In this text, most numbers are assumed to have three significant figures. Furthermore, consistent numbers of significant figures are used in all worked examples. An answer given to three digits is based on input good to at least three digits, for example. If the input has fewer significant figures, the answer will also have fewer significant figures. Care is also taken that the number of significant figures is reasonable for the situation posed. In some topics, particularly in optics, more accurate numbers are needed and we use more than three significant figures. Finally, if a number is **exact**, such as the two in the formula for the circumference of a circle, $C = 2\pi r$, it does not affect the number of significant figures in a calculation. Likewise, conversion factors such as 100 cm/1 m are considered exact and do not affect the number of significant figures in a calculation.

Chapter 2

2.0 Objectives

At the end of this lesson, students should be able to,

1. Identify displacement, and distance.
2. Calculate displacement and distance.
3. Distinguish between scalar and vector quantities.
4. Identify the relationships between instantaneous velocity, average velocity, instantaneous speed, average speed, displacement, and time.
5. Calculate velocity and speed.
6. Derive a graph of velocity vs. time given a graph of position vs. time.
7. Identify and distinguish between instantaneous acceleration, average acceleration, and deceleration.
8. Calculate acceleration.
9. Solve problems of one-dimensional kinematics.
10. Identify the motion of objects that are in free fall.
11. Calculate the position and velocity of objects in free fall.

2.1 Introduction

This chapter covers the topics related to one-dimensional kinematics including scalar and vector quantities, speed, velocity, acceleration, and free fall.

Kinematics is defined as the *study of motion without considering its causes.*

2.2 Position

In order to describe the motion of an object, you must first be able to describe its position—where it is at any particular time. More precisely, you need to specify its position relative to a convenient reference frame. Earth is often used as a reference frame, and we often describe the position of an object as it relates to stationary objects in that reference frame. For example, a rocket launch would be described in terms of the position of the rocket with respect to the Earth as a whole. In other cases, we use reference frames that are not stationary but are in motion relative to the Earth. To describe the position of a person in an airplane, for example, we use the airplane, not the Earth, as the reference frame.

2.3 Displacement

If an object moves relative to a reference frame (for example, if a professor moves to the right relative to a white board in a class or a passenger moves toward the rear of an airplane), then the object's position changes. This change in position is known as displacement. The word "displacement" implies that an object has moved, or has been displaced.

Displacement is the *change in position* of an object:

$$\Delta x = x_f - x_0 \qquad 2.1$$

where Δx is displacement, x_f is the final position, and x_0 is the initial position.

In this text the uppercase Greek letter Δ (delta) always means "change in" whatever quantity follows it; thus, Δx means **change in position**. Always solve for displacement by subtracting initial position x_0 from final position x_f.

Note that the SI unit for displacement is the meter (m), but sometimes kilometers, miles, feet, and other units of length are used. Keep in mind that when units other than the meter are used in a problem, you may need to convert them into meters to complete the calculation.

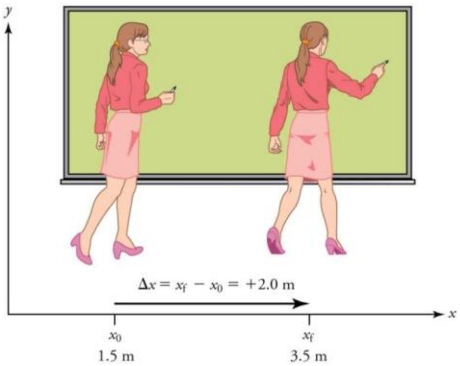


Figure 2.1: A professor paces left and right while lecturing. Her position relative to Earth is given by *x*. The +2.0m displacement of the professor relative to Earth is represented by an arrow pointing to the right.

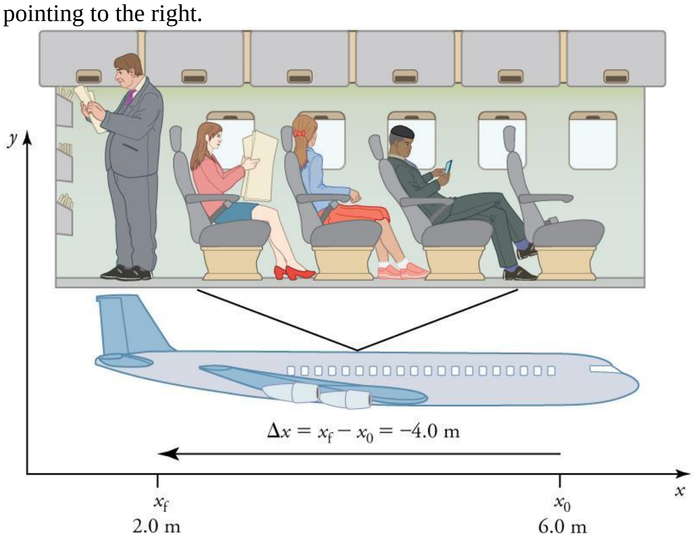


Figure 2.2: A passenger moves from his seat to the back of the plane. His location relative to the airplane is given by x. The −4.0 m displacement of the passenger relative to the plane is represented by an arrow toward the rear of the plane. Notice that the arrow representing his displacement is twice as long as the arrow representing the displacement of the professor (he moves twice as far).

Note that displacement has a direction as well as a magnitude. The professor's displacement is 2.0 m to the right, and the airline passenger's displacement is 4.0 m toward the rear. In one-dimensional motion, direction can be specified with a plus or minus sign. When you begin a problem, you should select which direction is positive (usually that will be to the right or up, but you are free to select positive as being any direction; nevertheless, you must stick to the same selection throughout a problem). The professor's initial position is x_0 =1.5m and her final position is x_f = 3.5m. Thus, her displacement is,

$$\Delta x = x_f - x_0 = 3.5\text{ m} - 1.5\text{ m} = +2.0\text{ m} \qquad 2.2$$

In this coordinate system, motion to the right is positive, whereas motion to the left is negative. Similarly, the airplane passenger's initial position is x_0 = 6.0 m and his final position is x_f = 2.0 m, so his displacement is

$$\Delta x = x_f - x_0 = 2.0\text{ m} - 6.0\text{ m} = -4.0\text{ m} \qquad 2.3$$

His displacement is negative because his motion is toward the rear of the plane, or in the negative x direction in our coordinate system.

2.4 Distance

Although displacement is described in terms of direction, distance is not. Distance is defined to be **the magnitude or size of displacement between two positions**.

Misconception Alert: Distance Traveled vs. Magnitude of Displacement

It is important to note that the *distance traveled*, however, can be greater than the magnitude of the displacement (by magnitude, we mean just the size of the displacement without regard to its direction; that is, just a number with a unit). For example, the professor could pace back and forth many times, perhaps walking a distance of 150 m during a lecture, yet still end up only 2.0 m to the right of her starting point. In this case her displacement would be +2.0 m, the magnitude of her displacement would be 2.0 m, but the distance she traveled would be 150 m. In kinematics we nearly always deal with displacement and magnitude of displacement, and almost never with distance traveled. One way to think about this is to assume you marked the start of the motion and the end of the motion. The displacement is simply the difference in the position of the two marks and is independent of the path taken in traveling between the two marks. The distance traveled, however, is the total length of the path taken between the two marks.

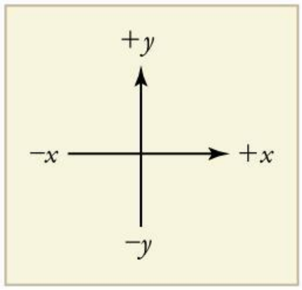
Displacement is defined by both direction and magnitude; distance is defined only by magnitude. Displacement is an example of a vector quantity. Distance is an example of a scalar quantity. A vector is any quantity with both *magnitude and direction*. Examples of vectors include a velocity of 90 km/h east and a force of 500 newtons straight down.

The direction of a vector in one-dimensional motion is given simply by a plus (+) or minus (−) sign. Vectors are represented graphically by arrows. An arrow used to represent a vector has a length proportional to the vector's magnitude (e.g., the larger the magnitude, the longer the length of the vector) and points in the same direction as the vector.

Some physical quantities, like distance, either have no direction or none is specified. A scalar is any quantity that has a magnitude, but no direction. For example, a 20ºC temperature, the 250 kilocalories (250 Calories) of energy in a candy bar, a 90 km/h speed limit, a person's 1.8 m height, and a distance of 2.0 m are all scalars — quantities with no specified direction. Note, however, that a scalar can be negative, such as a −20ºC temperature. In this case, the minus sign indicates a point on a scale rather than a direction. Scalars are never represented by arrows.

Coordinate Systems for One-Dimensional Motion

In order to describe the direction of a vector quantity, you must designate a coordinate system within the reference frame. For one-dimensional motion, this is a simple coordinate system consisting of a one-dimensional coordinate line. In general, when describing horizontal motion, motion to the right is usually considered positive, and motion to the left is considered negative. With vertical motion, motion up is usually positive, and motion down is negative. In some cases, however, it can be more convenient to switch the positive and negative directions. For example, if you are analyzing the motion of falling objects, it can be useful to define downwards as the positive direction. If people in a race are running to the left, it is useful to define left as the positive direction. **It does not matter as long as the system is clear and consistent. Once you assign a positive direction and start solving a problem, you cannot change it.**



Graph 2.1: It is usually convenient to consider motion upward or to the right as positive (+) and motion downward or to the left as negative (−).

2.5 Time

As discussed previously, the most fundamental physical quantities are defined by how they are measured. This is the case with time. Every measurement of time involves measuring a change in some physical quantity. It may be a number on a digital clock, a heartbeat, or the position of the Sun in the sky. In physics, the definition of time is simple — time is change, or the interval over which change occurs. It is impossible to know that time has passed unless something changes.

The amount of time or change is calibrated by comparison with a standard. The SI unit for time is the second, abbreviated s. We might, for example, observe that a certain pendulum makes one full swing every 0.75 s. We could then use the pendulum to measure time by counting its swings or, of course, by connecting the pendulum to a clock mechanism that registers time on a dial. This allows us to not only measure the amount of time, but also to determine a sequence of events.

How does time relate to motion? We are usually interested in elapsed time for a particular motion, such as how long it takes an airplane passenger to get from his seat to the back of the plane. To find elapsed time, we note the time at the beginning and end of the motion and subtract the two. For example, a lecture may start at 11:00 A.M. and end at 11:50 A.M., so that the elapsed time would be 50 min. Elapsed time Δt is the difference between the ending time and beginning time,

$$\Delta t = t_f - t_0 \qquad \textbf{2.4}$$

where Δt is the change in time or elapsed time, t_f is the time at the end of the motion, and t_0 is the time at the beginning of the motion.

Life is simpler if the beginning time t_0 is taken to be zero, as when we use a stopwatch. If we were using a stopwatch, it would simply read zero at the start of the lecture and 50 min at the end. If $t_0 = 0$, then $\Delta t = t_f = t$.

In this text, for simplicity's sake,

- motion starts at time equal to zero ($t_0 = 0$)
- the symbol t is used for elapsed time unless otherwise specified ($\Delta t = t_f = t$)

2.6 Velocity

Velocity is the displacement per unit time. You know that if you have a large displacement in a small amount of time you have a large velocity, and that velocity has units of displacement divided by time, such as miles per hour or kilometers per hour.

Velocity = Displacement / Time

Average Velocity

Average velocity is **displacement (change in position ($X_f - X_0$) divided by the time of travel**.

$$\mathbf{v_{av} = \Delta x/\Delta t = (x_f - x_0) / (t_f - t_0)} \qquad \mathbf{2.5}$$

where v_{av} is the *average* velocity, Δx is the change in position (or displacement), and x_f and x_0 are the final and beginning positions at times t_f and t_0, respectively. If the starting time t_0 is taken to be zero, then the average velocity is simply

$$\mathbf{v_{av} = \Delta x/t} \qquad \mathbf{2.6}$$

When the acceleration is constant, the average velocity is calculated by adding the initial velocity (V_0) and the final velocity (V_f) and then dividing the total by 2.

Notice that this definition indicates that *velocity is a vector because displacement is a vector*. It has both magnitude and direction. The SI unit for velocity is meters per second or m/s, but many other units, such as km/h, mi/h (also written as mph), and cm/s, are in common use. Suppose, for example, an airplane passenger took 5 seconds to move −4 m (the negative sign indicates that displacement is toward the back of the plane). His average velocity would be

$$\mathbf{v_{av} = \Delta x/t = -4m/5s = -0.8\ m/s}$$

The minus sign indicates the average velocity is also toward the rear of the plane.

The average velocity of an object does not tell us anything about what happens to it between the starting point and ending point. For example, we cannot tell from average velocity whether the airplane passenger stops momentarily or backs up before he goes to the back of the plane. To get more details, we must consider smaller segments of the trip over smaller time intervals.

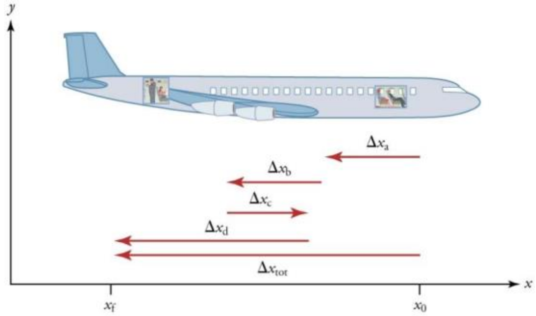


Figure 2.3: A more detailed record of an airplane passenger heading toward the back of the plane, showing smaller segments of his trip.

The smaller the time intervals considered in a motion, the more detailed the information. When we carry this process to its logical conclusion, we are left with an infinitesimally small interval. Over such an interval, the average velocity becomes the instantaneous velocity or the velocity at a specific instant. A car's speedometer, for example, shows the magnitude (but not the direction) of the instantaneous velocity of the car. (Police give tickets based on instantaneous velocity, but when calculating how long it will take to get from one place to another on a road trip, you need to use average velocity.) Instantaneous velocity v is the average velocity at a specific instant in time (or over an infinitesimally small-time interval).

2.7 Speed

In everyday language, most people use the terms "speed" and "velocity" interchangeably. In physics, however, they do not have the same meaning and they are distinct concepts. One major difference is that speed has no direction. Thus *speed is a scalar*. Just as we need to distinguish between instantaneous velocity and average velocity, we also need to distinguish between instantaneous speed and average speed.

Instantaneous speed is the magnitude of instantaneous velocity. For example, suppose the airplane passenger at one instant had an instantaneous velocity of −3.0 m/s (the minus meaning toward the rear of the plane), his instantaneous speed was 3.0 m/s – expressed without the direction. Or suppose that at one time during a shopping trip your instantaneous velocity is 40 km/h due north, then the instantaneous speed at that instant would be 40 km/h—the same magnitude but without a direction. Average speed, however, is very different from average velocity. Average speed is the distance traveled divided by elapsed time.

We have noted that distance traveled can be greater than the magnitude of displacement. So average speed can be greater than average velocity, which is displacement divided by time. For example, if you drive to a store and return home in half an hour, and your car's odometer shows the total distance traveled was 6 km, then your average speed was 12 km/h. Your average velocity, however, was zero, because your displacement for the round trip is zero. (Displacement is change in position and, thus, is zero for a round trip.) Thus, average speed is *not* simply the magnitude of average velocity.

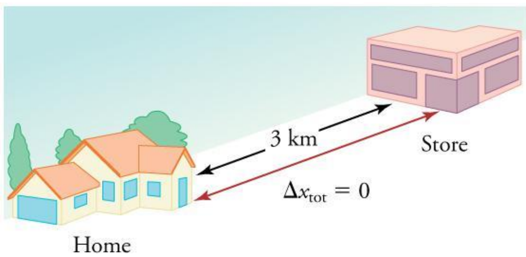
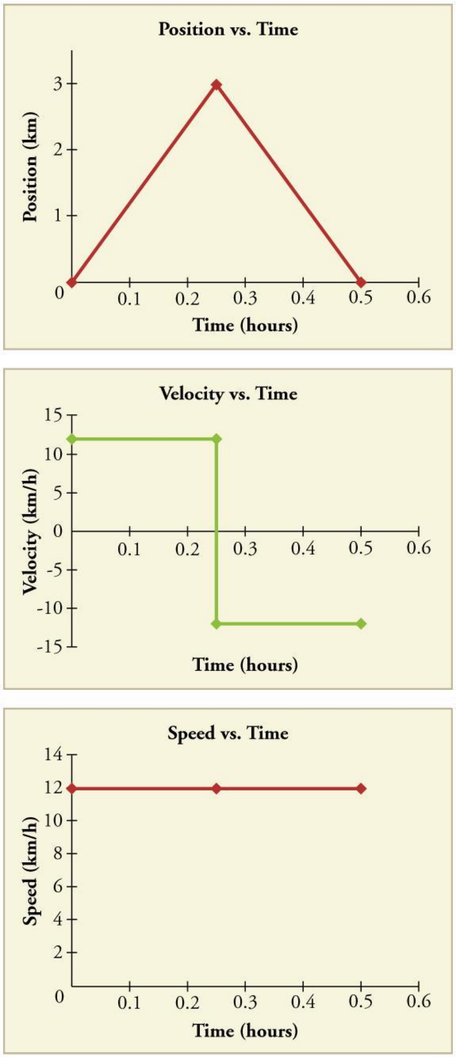


Figure 2.4: During a 30-minute round trip to the store, the total distance traveled is 6 km. The average speed is 12 km/h. The displacement for the round trip is zero, since there was no net change in position; therefore, the average velocity is zero.

Another way of visualizing the motion of an object is to use a graph. A plot of position or of velocity as a function of time can be very useful. For example, for this trip to the store, the position, velocity, and speed-vs.-time graphs are displayed in Graph 2.2. (Note that these graphs depict a very simplified model of the trip. We are assuming that speed is constant during the trip, which is unrealistic given that we'll probably stop at the store. But for simplicity's sake, we will model it with no stops or changes in speed. We are also assuming that the route between the store and the house is a perfectly straight line.)



Graph 2.2: Position vs. time, velocity vs. time, and speed vs. time on a trip. Note that the velocity for the return trip is negative.

2.8 Acceleration

Average Acceleration

Average Acceleration is *the rate at which velocity changes,*

$$\boldsymbol{a_{av} = \Delta v/\Delta t = (v_f - v_0)/(t_f - t_0)} \qquad \textbf{2.7}$$

where a_{av} is average acceleration, v is velocity, and t is time.

Because acceleration is velocity in m/s divided by time in s, the SI units for acceleration are m/s^2, meters per second squared or meters per second per second, which literally means by how many meters per second the velocity changes every second.

Recall that velocity is a vector—it has both magnitude and direction. This means that a change in velocity can be a change in magnitude (or speed), but it can also be a change in *direction*. For example, if a car turns a corner at constant speed, it is accelerating because its direction is changing. The quicker you turn, the greater the acceleration. So, there is an acceleration when velocity changes either in magnitude (an increase or decrease in speed) or in direction, or both.

Acceleration is a vector in the same direction as the *change* in velocity, Δv. Since velocity is a vector, it can change either in magnitude or in direction. Acceleration is therefore a change in either speed or direction, or both.

Keep in mind that although acceleration is in the direction of the *change* in velocity, it is not always in the direction of *motion*. When an object slows down, its acceleration is opposite to the direction of its motion. This is known as deceleration.

Important Notes: Deceleration vs. Negative Acceleration

Deceleration always refers to acceleration in the direction opposite to the direction of the velocity. **Deceleration always reduces speed. Negative acceleration, however, is acceleration** *in the negative direction in the chosen coordinate system***.** Negative acceleration may or may not be deceleration, and deceleration may or may not be considered negative acceleration.

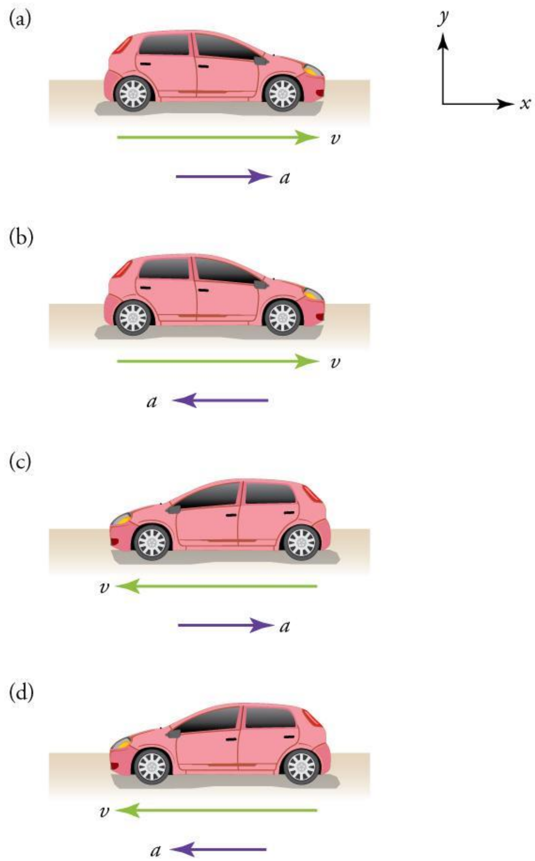


Figure **2.5:** (a) This car is speeding up as it moves toward the right. It therefore has positive acceleration in our coordinate system. (b) This car is slowing down as it moves toward the right. Therefore, it has negative acceleration in our coordinate system, because its acceleration is toward the left. The car is also decelerating: the direction of its acceleration is opposite to its direction of motion. (c) This car is moving toward the left, but slowing down over time. Therefore, its acceleration is positive in our coordinate system because it is toward the right. However, the car is decelerating because its acceleration is opposite to its motion. (d) This car is speeding up as it moves toward the left. It has negative acceleration because it is accelerating toward the left. However, because its acceleration is in the same direction as its motion, it is speeding up (*not* decelerating).

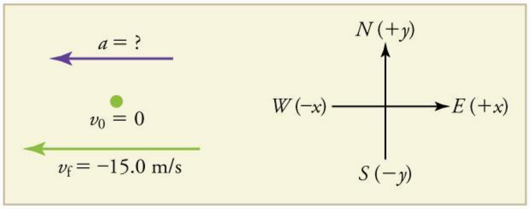
Example:

A racehorse coming out of the gate accelerates from rest to a velocity of 15.0 m/s due west in 1.80 s. What is its average acceleration?

Figure **2.6: (credit: Jon Sullivan, PD Photo.org)**

Strategy

Firstly, we draw a sketch and assign a coordinate system to the problem. This is a simple problem, but it always helps to visualize it. Notice that we assign east as positive and west as negative. Thus, in this case, we have negative velocity.



We can solve this problem by identifying Δv and Δt from the given information and then calculating the average acceleration directly from the equation $a_{av} = (\Delta v/\Delta t) = (v_f - v_0) / (t_f - t_0)$

Solution

1. Identify the knowns. $v_0 = 0$, $v_f = -15.0$ m/s (the negative sign indicates direction toward the west), $\Delta t = 1.80$ s.

2. Find the change in velocity. Since the horse is going from zero to −15.0 m/s, its change in velocity equals its final velocity: $\Delta v = v_f = -15.0$ m/s.

3. Plug in the known values (Δv and Δt) and solve for the unknown a_{av}.

$$a_{av} = \Delta v/\Delta t = -15.0 \text{ m/s} / 1.80 \text{ s} = -8.33 \text{ m/s}^2$$

Discussion

The negative sign for acceleration indicates that acceleration is toward the west. An acceleration of 8.33 m/s^2 due west means that the horse increases its velocity by 8.33 m/s due west each second, that is, 8.33 meters per second per second, which we write as 8.33 m/s^2. This is truly an average acceleration, because the ride is not smooth. We shall see later that an acceleration of this magnitude would require the rider to hang on with a force nearly equal to his weight.

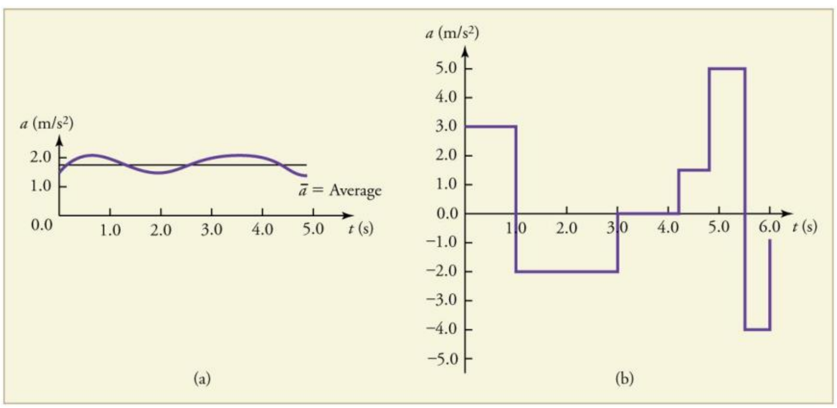
Instantaneous Acceleration

Instantaneous acceleration a, or the *acceleration at a specific instant in time*, is obtained by the same process as discussed for instantaneous velocity in Time, Velocity, and Speed - that is, by considering an infinitesimally small interval of time. How do we find instantaneous acceleration using only algebra? The answer is that we choose an average acceleration that is representative of the motion.

Graph 2.3 shows graphs of instantaneous acceleration versus time for two very different motions.

In graph 2.3 (a), the acceleration varies slightly and the average over the entire interval is nearly the same as the instantaneous acceleration at any time. In this case, we should treat this motion as if it had a constant acceleration equal to the average (in this case about 1.8 m/s^2).

In graph 2.3 (b), the acceleration varies drastically over time. In such situations it is best to consider smaller time intervals and choose an average acceleration for each. For example, we could consider motion over the time intervals from 0 to 1.0 s and from 1.0 to 3.0 s as separate motions with accelerations of +3.0 m/s^2 and –2.0 m/s^2, respectively.



Graph 2.3: Graphs of instantaneous acceleration versus time for two different one-dimensional motions. (a) Here acceleration varies only slightly and is always in the same direction, since it is positive. The average over the interval is nearly the same as the acceleration at any given time. (b) Here the acceleration varies greatly, perhaps representing a package on a post office conveyor belt that is accelerated forward and backward as it bumps along. It is necessary to

consider small time intervals (such as from 0 to 1.0 s) with constant or nearly constant acceleration in such a situation.

Examples on solving problems related to this chapter:

The examples below consider the motion of the subway train shown in Figure 2.10.

In (a) the shuttle moves to the right, and in (b) it moves to the left. The examples are designed to further illustrate aspects of motion and to illustrate some of the reasoning that goes into solving problems.

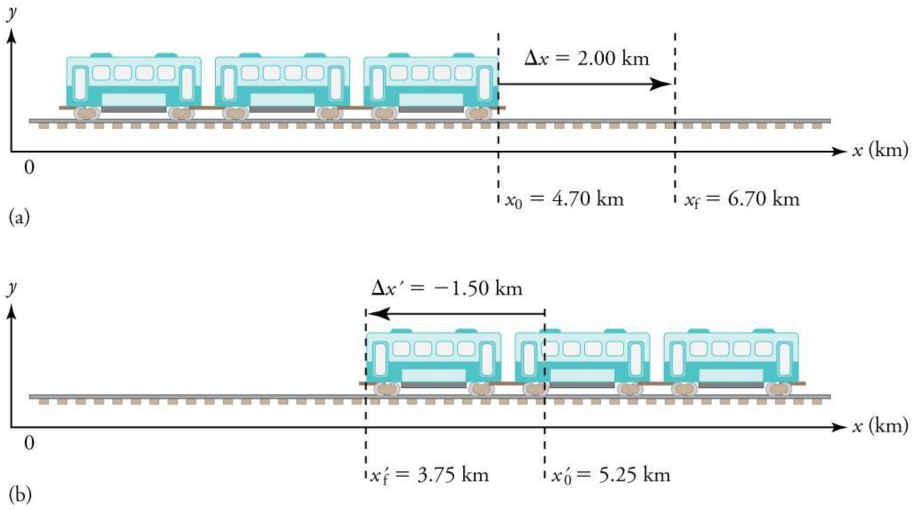


Figure **2.7:** One-dimensional motion of a subway train considered is shown above. Here we have chosen the x−axis so that + means to the right and − means to the left for displacements, velocities, and accelerations.

Example 1: Calculating Displacement

(a) The subway train moves to the right from $x0$ to x_f. Its displacement Δx is +2.0 km.

(b) The train moves to the left from $x'0$ to x'_f. Its displacement $\Delta x'$ is −1.5 km. (Note that the prime symbol (′) is used simply to distinguish between displacement in the two different situations. The distances of travel and the size of the cars are on different scales to fit everything into the diagram.)

What are the magnitude and sign of displacements for the motions of the subway train shown in parts (a) and (b) of Figure 2.10?

Strategy

A drawing with a coordinate system is already provided, so we don't need to make a sketch, but we should analyze it to make sure we understand what it is showing. Pay particular attention to the coordinate system. To find displacement, we use the equation $\Delta x = x_f - x_0$. This is straightforward since the initial and final positions are given.

Solution

1. Identify the knowns. In the figure we see that $x_f = 6.70$ km and $x_0 = 4.70$ km for part (a), and $x'_f = 3.75$ km and $x'_0 = 5.25$ km for part (b).

2. Solve for displacement in part (a).

$$\Delta x = x_f - x_0 = 6.70 \text{ km} - 4.70 \text{ km} = +2.00 \text{ km}$$

3. Solve for displacement in part (b).

$$\Delta x' = x'_f - x'_0 = 3.75 \text{ km} - 5.25 \text{ km} = -1.50 \text{ km}$$

Discussion

The direction of the motion in (a) is to the right and therefore its displacement has a positive sign, whereas motion in (b) is to the left and thus has a negative sign.

Example 2 - Comparing Distance Traveled with Displacement

What are the distances traveled for the motions shown in parts (a) and (b) of the subway train in Figure 2.10?

Strategy

To answer this question, think about the definitions of distance and distance traveled, and how they are related to displacement. Distance between two positions is defined to be the magnitude of displacement, which was found in example 2. Distance traveled is the total length of the path traveled between the two positions. In the case of the subway train, the distance traveled is the same as the distance between the initial and final positions of the train.

Solution

1. The displacement for part (a) was +2.00 km. Therefore, the distance between the initial and final positions was 2.00 km, and the distance traveled was 2.00 km.

2. The displacement for part (b) was −1.5 km. Therefore, the distance between the initial and final positions was 1.50 km, and the distance traveled was 1.50 km.

Discussion

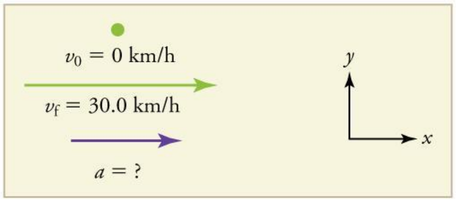
Distance is a scalar. It has magnitude but no sign to indicate direction.

Example 3 - Calculating Acceleration: A Subway Train Speeding Up

Suppose the train accelerates from rest to 30.0 km/h in the first 20.0 s of its motion. What is its average acceleration during that time interval?

Strategy

It is worth it at this point to make a simple sketch:



This problem involves three steps. First, we must determine the change in velocity, then we must determine the change in time, and finally we use these values to calculate the acceleration.

Solution

1. Identify the knowns. v_0=0 (the trains starts at rest), v_f = 30.0 km/h, and Δt = 20.0 s.

2. Calculate Δv. Since the train starts from rest, its change in velocity is Δv = +30.0 km/h, where the plus sign means velocity to the right.

3. Plug in known values and solve for the unknown, a_{av}.

4. Since the units are mixed (we have both hours and seconds for time), we need to convert everything into SI units of meters and seconds.

$$a_{av} = ((+30 \text{ km/h}) / 20.0 \text{ s}) \text{ X } (10^3\text{m} / 1 \text{ km}) \text{ X } (1 \text{ h} / 3600 \text{ s}) = 0.417 \text{ m/s}^2$$

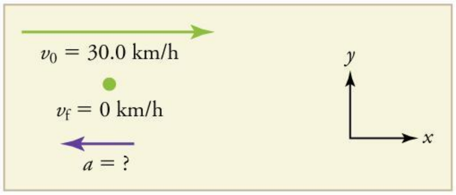
Discussion

The plus sign means that acceleration is to the right. This is reasonable because the train starts from rest and ends up with a velocity to the right (also positive). So, acceleration is in the same direction as the *change* in velocity, as is always the case.

Example 4 - Calculate Acceleration: A Subway Train Slowing Down

Now suppose that at the end of its trip, the train slows to a stop from a speed of 30.0 km/h in 8.00 s. What is its average acceleration while stopping?

Strategy



In this case, the train is decelerating and its acceleration is negative because it is toward the left. As in the previous example, we must find the change in velocity and the change in time and then solve for acceleration.

Solution

1. Identify the knowns. v_0=30.0 km/h, v_f = 0 km/h (the train is stopped, so its velocity is 0), and Δt = 8.00 s.

2. Solve for the change in velocity, Δv.

$$\Delta v = v_f - v_0 = 0 - 30.0 \text{ km/h} = -30.0 \text{ km/h}$$

3. Plug in the knowns, Δv and Δt, and solve for a_{av}.

$$a_{av} = \Delta v/\Delta t = (-30.0 \text{ km/h}) /8.00 \text{ s}$$

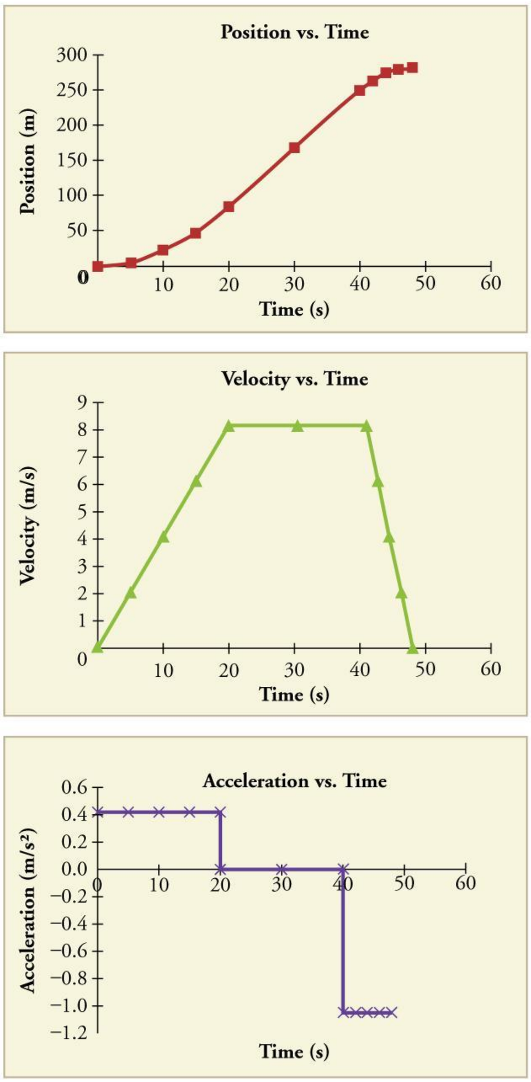
4. Convert the units to meters and seconds.

$$a_{av} = \Delta v/\Delta t = ((-30.0 \text{ km/h}) / 8.00 \text{ s}) \text{ X } (10^3 \text{m} / 1 \text{ km}) \text{ X } (1 \text{ h}/ 3600 \text{ s}) = -1.04 \text{ m/s}^2.$$

Discussion

The minus sign indicates that acceleration is to the left. This sign is reasonable because the train initially has a positive velocity in this problem, and a negative acceleration would oppose the motion. Again, acceleration is in the same direction as the *change* in velocity, which is negative here. This acceleration can be called a deceleration because it has a direction opposite to the velocity.

The graphs of position, velocity, and acceleration vs. time for the trains in above examples are displayed in graph 2.4. We have taken the velocity to remain constant from 20 to 40 s, after which the train decelerates.



Graph 2.4: (a) Position of the train overtime. Notice that the train's position changes slowly at the beginning of the journey, then more and more quickly as it picks up speed. Its position then changes more slowly as it slows down at the end of the journey. In the middle of the journey, while the velocity remains constant, the position changes at a constant rate. (b) Velocity of the train over time. The train's velocity increases as it accelerates at the beginning of the journey. It remains the same in the middle of the journey (where there is no acceleration). It decreases as the train decelerates at the end of the journey. (c) The acceleration of the train over time. The train

has positive acceleration as it speeds up at the beginning of the journey. It has no acceleration as it travels at constant velocity in the middle of the journey. Its acceleration is negative as it slows down at the end of the journey.

Example 5 - Calculating Average Velocity: The Subway Train

What is the average velocity of the train in part b of example 1, and shown again below, if it takes 5.00 min to make its trip?

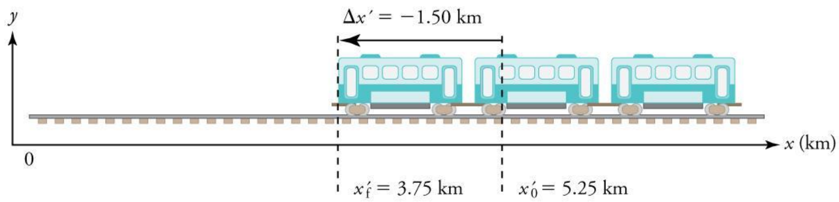


Figure **2.8**

Strategy
Average velocity is displacement divided by time. It will be negative here, since the train moves to the left and has a negative displacement.

Solution
1. Identify the knowns. $x'_f = 3.75$ km, $x'_0 = 5.25$ km, $\Delta t = 5.00$ min.

2. Determine displacement, $\Delta x'$. We found $\Delta x'$ to be −1.5 km in example 1.

3. Solve for average velocity.

$$v_{av} = \Delta x'/\Delta t = -1.50 \text{ km} / 5.00 \text{ min}$$

4. Convert units.

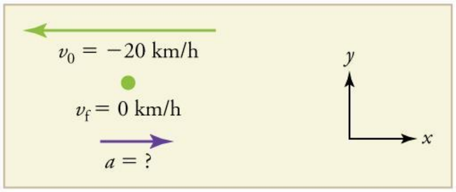
$$v_{\text{av}} = \Delta x'/\Delta t = (-1.50 \text{ km} / 5.00 \text{ min}) \text{ X } (60 \text{ min}/ 1 \text{ h}) = -18.0 \text{ km/h}$$

Discussion
The negative velocity indicates motion to the left.

Example 6 - Calculating Deceleration: The Subway Train

Finally, suppose the train in Figure 2.12 slows to a stop from a velocity of 20.0 km/h in 10.0 s. What is its average acceleration?

Strategy
Once again, let’s draw a sketch:



As before, we must find the change in velocity and the change in time to calculate average acceleration.

Solution

1. Identify the knowns. $v_0 = -20$ km/h, $v_f = 0$ km/h, $\Delta t = 10.0$ s.

2. Calculate Δv. The change in velocity here is actually positive, since

$$\Delta v = v_f - v_0 = 0 - (-20\text{ km/h}) = +20\text{ km/h}.$$

3. Solve for a_{av}.

$$a_{av} = \Delta v/\Delta t = +20.0\text{ km/h}/10.0\text{ s}$$

4. Convert units.

$$a_{av} = ((+20.0\text{ km/h}) / 10.0\text{ s})\ X\ (10^3\text{m} / 1\text{ km})\ X\ (1\text{ h} / 3600\text{ s}) = +\ 0.556\text{ m/s2}$$

Discussion

The plus sign means that acceleration is to the right. This is reasonable because the train initially has a negative velocity (to the left) in this problem and a positive acceleration opposes the motion (and so it is to the right). Again, acceleration is in the same direction as the *change* in velocity, which is positive here. This acceleration can be called a deceleration since it is in the direction opposite to the velocity.

Sign and Direction

Perhaps the most important thing to note about these examples is the signs of the answers. In our chosen coordinate system, plus means the quantity is to the right and minus means it is to the left. This is easy to imagine for displacement and velocity. But it is a little less obvious for acceleration. Most people interpret negative acceleration as the slowing of an object. As discussed above, when a positive acceleration slowed a negative velocity, the crucial distinction was that the acceleration was in the opposite direction from the velocity. In fact, a negative acceleration will *increase* a negative velocity. For example, the train moving to the left in Figure 2.12 is sped up by an acceleration to the left. In that case, both v and a are negative. The plus and minus signs give the directions of the accelerations. If acceleration has the same sign as the velocity, the object is speeding up. If acceleration has the opposite sign as the velocity, the object is slowing down.

2.9 Kinematic Equations

Notations: *t, x, v, a*

Before we discuss the kinematic equations, let us make some simplifications in notations. We can consider initial time to be zero in most of the cases. Since elapsed time is $\Delta t = t_f - t_0$, taking $t_0=0$ means that $\Delta t = t_f$, the final time on the stopwatch. When initial time is taken to be zero, we use the subscript 0 to denote initial values of position and velocity. That is, x_0 *is the initial position* and v_0 *is the initial velocity*. We put no subscripts on the final values. That is, t *is the final time*, x *is the final position*, and v *is the final velocity*. This gives a simpler expression.

For elapsed time, $\Delta t = t$.

For displacement, which is now $\Delta x = x - x_0$.

For change in velocity, which is now $\Delta v = v - v_0$.

Note that *the subscript 0 denotes an initial value and the absence of a subscript denotes a final value* in whatever motion is under consideration.

We now make the important assumption that *acceleration is constant*. This assumption allows us to avoid using calculus to find instantaneous acceleration. Since acceleration is constant, the average and instantaneous accelerations are equal. That is,

$$a_{av} = a = \text{constant},$$

Therefore, we use the symbol *a* for acceleration at all times. Assuming acceleration to be constant does not seriously limit the situations we can study nor degrade the accuracy of our treatment. For one thing, acceleration *is* constant in a great number of situations. Furthermore, in many other situations we can accurately describe motion by assuming a constant acceleration equal to the average acceleration for that motion. Finally, in motions where acceleration changes drastically, such as a car accelerating to top speed and then braking to a stop, the motion can be considered in separate parts, each of which has its own constant acceleration.

Since we have explained the distance, displacement, speed, velocity, and acceleration, and how their magnitudes and directions vary depending on the situations, you should be able to directly apply the following kinematics equations to solve problems.

- Equation 1: $\boldsymbol{v = v_0 + at}$
  This means if you have a constant acceleration, you can find the final velocity. This equation is used when the situation involves changing velocities with a constant acceleration.

- Equation 2: $\mathbf{\Delta x = [\,(v + v_0)/2]t}$
  When acceleration is not given but you are given the changing velocities and the displacement, you can use this equation.

- Equation 3: $\mathbf{\Delta x = v_0 t + {}^1/_2 a t^2}$

This means the displacement can be related to the initial velocity and a constant acceleration without having to find the final velocity. You can use this equation when final velocity is the only value not given.

The above equation can also be written as,

$$\mathbf{x = x_0 + v_0t + {}^1/_2at^2}$$

(Please note that $X\text{-}X_0$ is Δx.)

The above equation reduces to,

$$\mathbf{x = x_0 + v_0t \;\text{ or }\; \Delta x = v_0t,}$$

when the velocity is constant, the acceleration is zero.

- Equation 4: $\mathbf{v^2 = v_0^2 + 2a(\Delta x)}$
  This equation can be used when the time is not given.

2.10 Free Fall and Gravity

All falling objects in the absence of friction and air resistance fall under the same constant acceleration independent of their mass. The constant acceleration on these objects is known as gravity. The acceleration of gravity on the surface of Earth is 9.8 m/s^2. They all follow one-dimensional kinematic laws. If you throw an object directly upward (not at an angle), that object also follows one dimensional path along the y-axis. When dealing with free fall problems or objects thrown directly upwards, always remember that gravity acts in the negative direction (toward the ground) and the objects moving upward will have a positive velocity (remember that upward is positive and downward is negative when you assign directions).

Kinematics Equations can be used to analyze Free-Fall, but they have to be applied in the vertical direction and acceleration becomes g. Therefore, x can be replaced with y for all those kinematic equations.

Therefore, the modified Kinematic Equations can be written as follows (the + and – signs may vary depending on the given problem):

- $\boldsymbol{v = v_0 - gt}$
- $\mathbf{\Delta y = [\,(v + v_0)/2]t}$
- $\mathbf{\Delta y = v_0t - {}^1/_2gt^2}$
- $\mathbf{v^2 = v_0^2 - 2g(\Delta y)}$

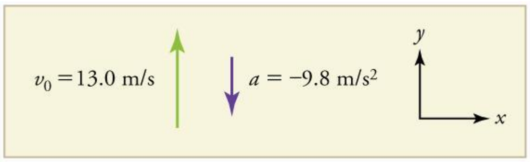
Please note that the above equations have already accounted for the fact that gravitational acceleration is in the negative direction. Therefore, you can directly apply the magnitude of gravitation accelerations (9.8 ms^2), without adding the negative sign.

Example of an application of kinematic equations to c**alculate position and velocity of a falling object: A Rock Thrown Upward**

A person standing on the edge of a high cliff throws a rock straight up with an initial velocity of 13.0 m/s. The rock misses the edge of the cliff as it falls back to earth. Calculate the position and velocity of the rock 1.00 s, 2.00 s, and 3.00 s after it is thrown, neglecting the effects of air resistance.

Strategy

Draw a sketch.



We are asked to determine the position y at various times. It is reasonable to take the initial position y_0 to be zero. This problem involves one-dimensional motion in the vertical direction. We use plus and minus signs to indicate direction, with up being positive and down negative. Since up is positive, and the rock is thrown upward, the initial velocity must be positive too. The acceleration due to gravity is downward, so a is negative. It is crucial that the initial velocity and the acceleration due to gravity have opposite signs. Opposite signs indicate that the acceleration due to gravity opposes the initial motion and will slow and eventually reverse it.

Since we are asked for values of position and velocity at three times, we will refer to these as y_1 and v_1; y_2 and v_2; and y_3 and v_3.

Solution for Position - y_1

1. Identify the knowns. We know that $y_0 = 0$; $v_0 = 13.0$ m/s; $a = -g = -9.80\ \text{m/s}^2$; and $t = 1.00$ s.

2. Identify the best equation to use. We will use $y = y_0 + v_0 t + ½\, at^2$ because it includes only one unknown, y (or y_1, here), which is the value we want to find.

3. Plug in the known values and solve for y_1.

$$y_1 = 0 + (13.0\ \text{m/s})(1.00\ \text{s}) + 12(-9.80\text{m/s}^2)(1.00\ \text{s})^2 = 8.10\text{m}$$

Discussion

The rock is 8.10 m above its starting point at $t = 1.00$ s, since $y_1 > y_0$. It could be *moving* up or down; the only way to tell is to calculate $v1$ and find out if it is positive or negative.

Solution for Velocity - v_1

1. Identify the knowns. We know that $y_0 = 0$; $v_0 = 13.0$ m/s; $a = -g = -9.80\ \text{m/s}^2$; and $t=1.00$ s. We also know from the solution above that $y_1=8.10$ m.

2. Identify the best equation to use. The most straightforward is $v = v_0 - gt$ (from $v = v_0+at$, where a = gravitational acceleration = $-g$).

3. Plug in the knowns and solve.

$$v_1 = v_0 - gt = 13.0\ \text{m/s} - (9.80\ \text{m/s}^2)(1.00\ \text{s}) = 3.20\ \text{m/s}$$

Discussion

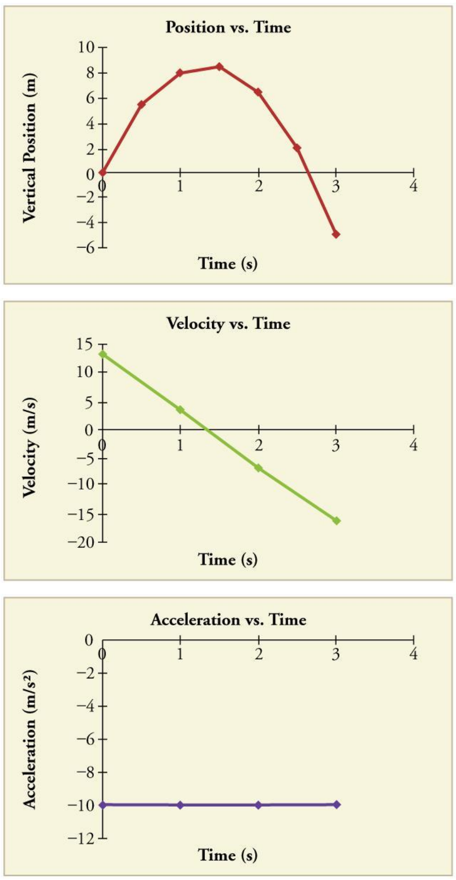
The positive value for v_1 means that the rock is still heading upward at t=1.00s. However, it has slowed from its original 13.0 m/s, as expected.

Solution for Remaining Times

The procedures for calculating the position and velocity at t=2.00s and 3.00 s are the same as those above. The results are given below:

Time, t	**Position,** y	**Velocity,** v	**Acceleration,** a
1.00 s	8.10 m	3.20 m/s	-9.80 m/s2
2.00 s	6.40 m	-6.60 m/s	-9.80 m/s2
3.00 s	-5.10 m	-16.4 m/s	-9.80 m/s2

Graphing the data helps us understand it more clearly.



Graph 2.5: Vertical position, vertical velocity, and vertical acceleration vs. time for a rock thrown vertically up at the edge of a cliff. Notice that velocity changes linearly with time and that acceleration is constant. *Misconception Alert!* Notice that the position vs. time graph shows vertical position only. It is easy to get the impression that the graph shows some horizontal motion—the shape of the graph looks like the path of a projectile. But this is not the case; the horizontal axis is *time*, not space. The actual path of the rock in space is straight up, and straight down.

Discussion

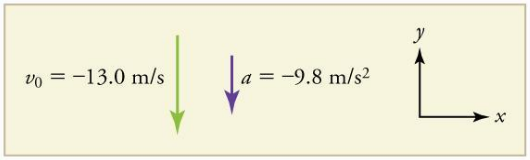
The interpretation of these results is important. At 1.00 s the rock is above its starting point and heading upward, since y_1 and v_1 are both positive. At 2.00 s, the rock is still above its starting point, but the negative velocity means it is moving downward. At 3.00 s, both y_3 and v_3 are negative, meaning the rock is below its starting point and continuing to move downward. Notice that when the rock is at its highest point (at 1.33 s), its velocity is zero, but its acceleration is still $-9.80\ m/s^2$. Its acceleration is $-9.80\ m/s^2$ for the whole trip—while it is moving up and while it is moving down. Note that the values for *y* are the positions (or displacements) of the rock, not the total distances traveled. Finally, note that free-fall applies to upward motion as well as downward. Both have the same acceleration—the acceleration due to gravity, which remains constant the entire time.

Example of an application of kinematic equations to c**alculate position and velocity of a falling object: A Rock Thrown Down**

What happens if the person on the cliff throws the rock straight down, instead of straight up? To explore this question, calculate the velocity of the rock when it is 5.10 m below the starting point, and has been thrown downward with an initial speed of 13.0 m/s.

Strategy

Draw a sketch.



Since up is positive, the final position of the rock will be negative because it finishes below the starting point at $y_0=0$. Similarly, the initial velocity is downward and therefore negative, as is the acceleration due to gravity. We expect the final velocity to be negative since the rock will continue to move downward.

Solution

1. Identify the knowns. $y_0 = 0$; $y_1 = -5.10\ m$; $v_0 = -13.0\ m/s$; $a = -g = -9.80\ m/s^2$.

2. Choose the kinematic equation that makes it easiest to solve the problem. The equation $v^2=v^2_0+2a(y-y_0)$ works well because the only unknown in it is *v*. (We will plug y_1 in for *y*.)

3. Enter the known values

$$v^2 = (-13.0\ m/s)^2 + 2(-9.80\ m/s^2)(-5.10\ m - 0\ m) = 268.96\ m^2/s^2,$$

where we have retained extra significant figures because this is an intermediate result.

Taking the square root, and noting that a square root can be positive or negative, gives

$$v = \pm 16.4\ m/s.$$

The negative root is chosen to indicate that the rock is still heading down. Thus,

$$v = -16.4\ m/s.$$

Discussion

Note that this is exactly the same velocity the rock had at this position when it was thrown straight upward with the same initial speed. This is not a coincidental result. Because we only consider the acceleration due to gravity in this problem, the speed of a falling object depends only on its initial speed and its vertical position relative to the starting point. For example, if the velocity of the rock is calculated at a height of 8.10 m above the starting point, when the initial velocity is 13.0 m/s straight up, a result of ±3.20 m/s is obtained. Here both signs are meaningful; the positive value occurs when the rock is at 8.10 m and heading up, and the negative value occurs when the rock is at 8.10 m and heading back down. It has the same speed but the opposite direction.

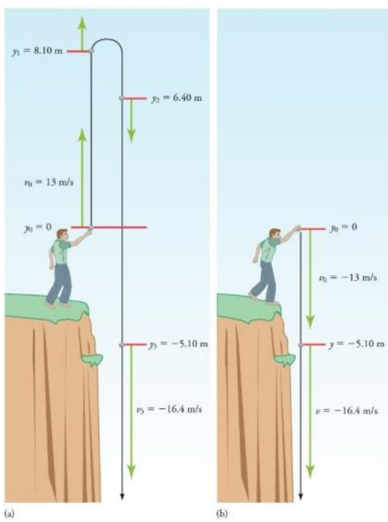
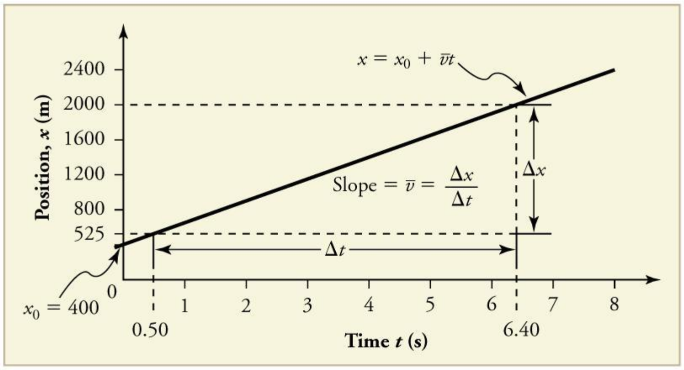


Figure 2.9: (a) A person throws a rock straight up, as explored in previously. The arrows are velocity vectors at 0, 1.00, 2.00, and 3.00 s. (b) A person throws a rock straight down from a cliff with the same initial speed as before. Note that at the same distance below the point of release, the rock has the same velocity in both cases. Another way to look at it is this: The rock is thrown up with an initial velocity of 13.0 m/s. It rises and then falls back down. When its position is $y=0$ on its way back down, its velocity is −13.0 m/s. That is, it has the same speed on its way down as on its way up. We would then expect its velocity at a position of $y=-5.10$ m to be the same whether we have thrown it upwards at +13.0 m/s or thrown it downwards at −13.0 m/s. The velocity of the rock on its way down from $y=0$ is the same whether we have thrown it up or down to start with, as long as the speed with which it was initially thrown is the same.

2.11 Graphs in Kinematics

Graph of Position vs. Time ($a = 0$, so v is constant)

Time is usually an independent variable that other quantities, such as position, depend upon. A graph of position versus time would, thus, have x on the vertical axis and t on the horizontal axis. Graph 2.6 is just such a straight-line graph. It shows a graph of position versus time for a jet-powered car.



Graph 2.6: Graph of position versus time for a jet-powered car.

Using the relationship between dependent and independent variables, we see that the slope in the graph above is average velocity and the intercept is position at time zero - that is, x_0. Substituting these symbols into $y = mx + b$ gives

$$x = v_{av}t + x_0$$

or

$$x = x_0 + v_{av}t.$$

Thus a graph of position versus time gives a general relationship among displacement(change in position), velocity, and time, as well as giving detailed numerical information about a specific situation.

The Slope of x vs. t

The slope of the graph of position x vs. time t is velocity v.

$$\text{Slope} = \Delta x/\Delta t = v$$

From the figure we can see that the car has a position of 525 m at 0.50 s and 2000 m at 6.40 s. Its position at other times can be read from the graph; furthermore, information about its velocity and acceleration can also be obtained from the graph.

Determining Average Velocity from a Graph of Position versus Time

Find the average velocity of the car whose position is graphed above.

Strategy

The slope of a graph of x vs. t is average velocity, since slope equals rise over run. In this case, rise = change in position and run = change in time, so that

$$\text{Slope} = \Delta x/\Delta t = v_{av}.$$

Since the slope is constant here, any two points on the graph can be used to find the slope. (Generally speaking, it is most accurate to use two widely separated points on the straight line. This is because any error in reading data from the graph is proportionally smaller if the interval is larger.)

Solution

1. Choose two points on the line. In this case, we choose the points labeled on the graph: (6.4 s, 2000 m) and (0.50 s, 525 m). (Note, however, that you could choose any two points.)

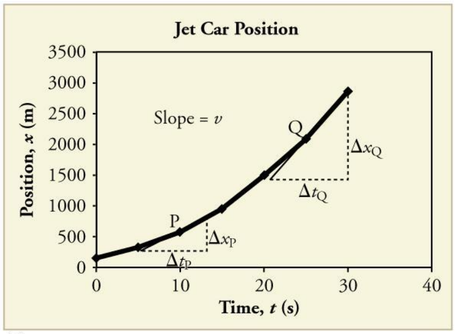
2. Substitute the x and t values of the chosen points into the equation. Remember in calculating change (Δ) we always use final value minus initial value.

$$v_{av} = \Delta x/\Delta t = (2000 \text{ m} - 525 \text{ m}) / (6.4 \text{ s} - 0.50 \text{ s}),$$

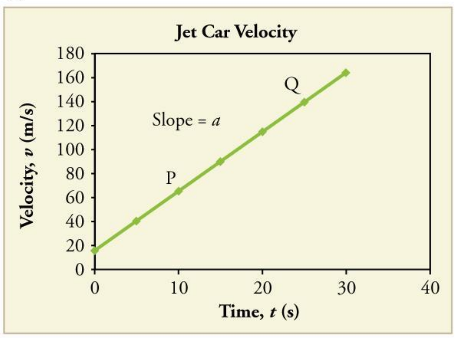
$$v_{av} = 250 \text{ m/s}.$$

Graphs of Motion when *a* is constant but $a \neq 0$

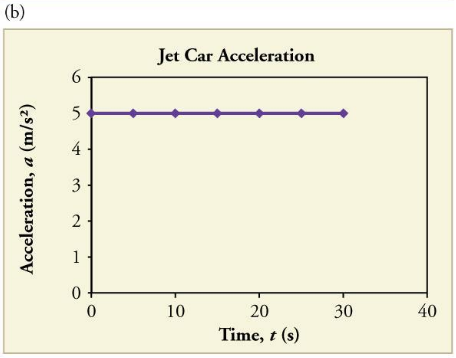
Graph 2.7 represents the motion of the jet-powered car as it accelerates toward its top speed, but only during the time when its acceleration is constant. Time starts at zero for this motion (as if measured with a stopwatch), and the position and velocity are initially 200 m and 15 m/s, respectively.



(a)



(b)



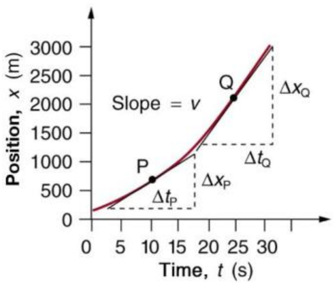
(c)

Graph **2.7:** Graphs of motion of a jet-powered car during the time span when its acceleration is constant. (a) The slope of an *x* vs. *t* graph is velocity. This is shown at two points, and the instantaneous velocities obtained are plotted in the next graph. Instantaneous velocity at any point is the slope of the tangent at that point. (b) The slope of the *v* vs. *t* graph is constant for this part of the motion, indicating constant acceleration. (c) Acceleration has the constant value of 5.0 m/s2 over the time interval plotted.

The graph of position versus time in graph 2.7 (a) is a curve rather than a straight line. The slope of the curve becomes steeper as time progresses, showing that the velocity is increasing over time. The slope at any point on a position-versus-time graph is the instantaneous velocity at that point. It is found by drawing a straight-line tangent to the curve at the point of interest and taking the slope of this straight line. Tangent lines are shown for two points in graph 2/7(a). If this is done at every point on the curve and the values are plotted against time, then the graph of velocity versus time shown in graph 2.7(b) is obtained. Furthermore, the slope of the graph of velocity versus time is acceleration, which is shown in graph 2.7(c).

Determining Instantaneous Velocity from the Slope at a Point: Jet Car

Calculate the velocity of the jet car at a time of 25 s by finding the slope of the *x* vs. *t* graph in the graph below.



t (s)	x (m)
0	200
5	338
10	600
15	988
20	1500
25	2138
30	2900

Graph **2.8:** The slope of an *x* vs. *t* graph is velocity. This is shown at two points. Instantaneous velocity at any point is the slope of the tangent at that point.

Strategy

The slope of a curve at a point is equal to the slope of a straight line tangent to the curve at that point. This principle is illustrated in Graph 2.8, where Q is the point at t=25 s.

Solution

1. Find the tangent line to the curve at t = 25 s.

2. Determine the endpoints of the tangent. These correspond to a position of 1300 m at time 19 s and a position of 3120 m at time 32 s.

3. Plug these endpoints into the equation to solve for the slope, v.

$$\text{Slope} = v_Q = \Delta x_Q/\Delta t_Q = (3120\text{ m} - 1300\text{ m})(32\text{ s} - 19\text{ s})$$

Thus,

$$v_Q = 1820\text{ m}/13\text{ s} = 140\text{ m/s}.$$

Discussion

This is the value given in this figure's table for v at t=25 s. The value of 140 m/s for v_Q is plotted in graph 2.8. The entire graph of v vs. t can be obtained in this fashion.

Carrying this one step further, we note that the slope of a velocity versus time graph is acceleration. Slope is rise divided by run; on a v vs. t graph, rise = change in velocity Δv and run = change in time Δt.

The Slope of *v* vs. *t*

The slope of a graph of velocity v vs. time t is acceleration a.

$$\text{Slope} = \Delta v/\Delta t = a$$

Since the velocity versus time graph in graph 2.8 (b) is a straight line, its slope is the same everywhere, implying that acceleration is constant.

Chapter 3

3.0 Objectives

At the end of this lesson, students should be able to,

1. Identify that motion in two dimensions consists of horizontal and vertical components.
2. Apply analytical methods to determine vertical and horizontal component vectors.
3. Identify and explain the properties of a projectile, such as acceleration due to gravity, range, maximum height, and trajectory.
4. Apply the principle of independence of motion to solve projectile motion problems.
5. Apply principles of vector addition to determine relative velocity.

3.1 Introduction

This chapter discusses two-dimensional kinematics. Both two- and three-dimensional kinematics are simple extensions of the one-dimensional kinematics developed for straight-line motion in the previous chapter. This simple extension will allow us to apply physics to many real-life situations.

3.2 Independence of Motion

The horizontal and vertical components of two-dimensional motion are independent of each other. Any motion in the horizontal direction does not affect motion in the vertical direction, and vice versa.

This is true in a simple scenario like that of walking in one direction first, followed by another. It is also true of more complicated motion involving movement in two directions at once. For example, let's compare the motions of two baseballs. One baseball is dropped from rest. At the same instant, another is thrown horizontally from the same height and follows a curved path. A stroboscope has captured the positions of the balls at fixed time intervals as they fall.

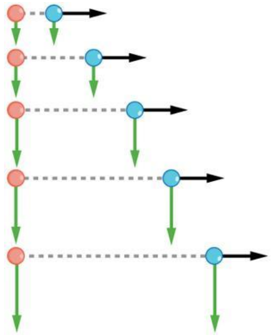


Figure 3.1: This figure shows the motions of two identical balls—one falls from rest, the other has an initial horizontal velocity. Each subsequent position is an equal time interval. Arrows represent horizontal and vertical velocities at each position. The ball on the right has an initial horizontal velocity, while the ball on the left has no horizontal velocity. Despite the difference in horizontal velocities, the vertical velocities and positions are identical for both balls. This shows that the vertical and horizontal motions are independent.

It is remarkable that for each flash of the strobe, the vertical positions of the two balls are the same. This similarity implies that the vertical motion is independent of whether or not the ball is moving horizontally. (Assuming no air resistance, the vertical motion of a falling object is influenced by gravity only, and not by any horizontal forces.) Careful examination of the ball thrown horizontally shows that it travels the same horizontal distance between flashes. This is due to the fact that there are no additional forces on the ball in the horizontal direction after it is thrown. This result means that the horizontal velocity is constant, and affected neither by vertical motion nor by gravity (which is vertical). Note that this case is true only for ideal conditions. In the real world, air resistance will affect the speed of the balls in both directions.

The two-dimensional curved path of the horizontally thrown ball is composed of two independent one-dimensional motions (horizontal and vertical). The key to analyzing such motion, called *projectile motion*, is to *resolve* (break) it into motions along perpendicular directions. Resolving two-dimensional motion into perpendicular components is possible because the components are independent.

3.2.1 Vectors in Two Dimensions

A vector is a quantity that has magnitude and direction. Displacement, velocity, acceleration, and force, for example, are all vectors. In one-dimensional, or straight-line, motion, the direction of a vector can be given simply by a plus or minus sign. In two dimensions (2-d), however, we specify the direction of a vector relative to some reference frame (i.e., coordinate system), using an arrow having length proportional to the vector’s magnitude and pointing in the direction of the vector.

Vectors in this Text

In this text, we will represent a vector with a boldface variable. For example, we will represent the quantity force with the vector $\mathbf{F}$, which has both magnitude and direction. The magnitude of the vector will be represented by a variable in italics, such as F, and the direction of the variable will be given by an angle ϑ.

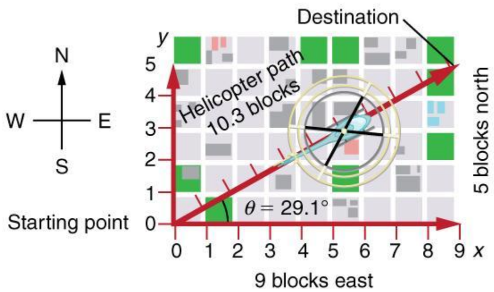


Figure 3.2: A person walks 9 blocks east and 5 blocks north. The displacement is 10.3 blocks at an angle 29.1º north of east (mentioned here as helicopter path taken directly).

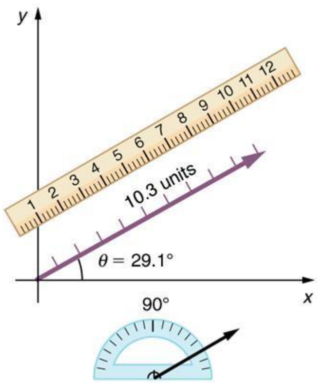


Figure 3.3: To describe the resultant vector for the person walking in a city considered in Figure 3.2 graphically, draw an arrow to represent the total displacement vector **D**. Using a protractor, draw a line at an angle θ relative to the east-west axis. The length D of the arrow is proportional to the vector's magnitude and is measured along the line with a ruler. In this example, the magnitude D of the vector is 10.3 units, and the direction θ is 29.1º north of east.

3.2.2 Vector Addition: Head-to-Tail Method

The head-to-tail method is a graphical way to add vectors, described in Figure 3.4 below and in the steps following. The tail of the vector is the starting point of the vector, and the head (or tip) of a vector is the final, pointed end of the arrow.

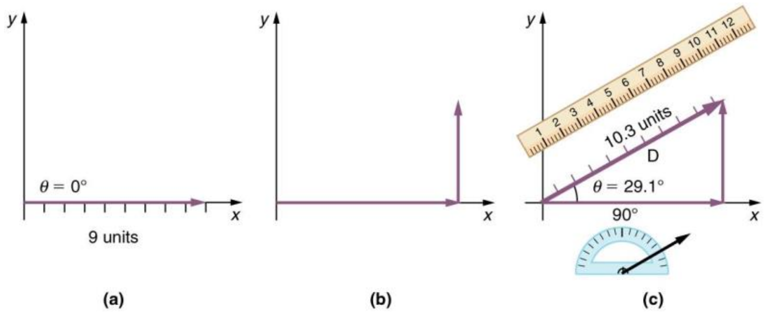


Figure 3.4: Head-to-Tail Method: The head-to-tail method of graphically adding vectors is illustrated for the two displacements of the person walking in a city considered in Figure 3.2 (a) Draw a vector representing the displacement to the east. (b) Draw a vector representing the displacement to the north. The tail of this vector should originate from the head of the first, east-pointing vector. (c) Draw a line from the tail of the east-pointing vector to the head of the north-pointing vector to form the sum or resultant vector **D**. The length of the arrow **D** is proportional to the vector's magnitude and is measured to be 10.3 units. Its direction, described as the angle with respect to the east (or horizontal axis) θ is measured with a protractor to be 29.1º.

Step 1. Draw an arrow to represent the first vector (9 blocks to the east) using a ruler and protractor.

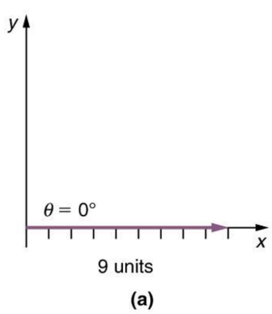


Figure 3.5:

Step 2. Now draw an arrow to represent the second vector (5 blocks to the north). *Place the tail of the second vector at the head of the first vector*.

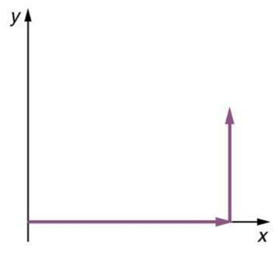


Figure 3.6:

Step 3. If there are more than two vectors, continue this process for each vector to be added. Note that in our example, we have only two vectors, so we have finished placing arrows tip to tail.

Step 4. Draw an arrow from the tail of the first vector to the head of the last vector. This is the resultant, or the sum, of the other vectors.

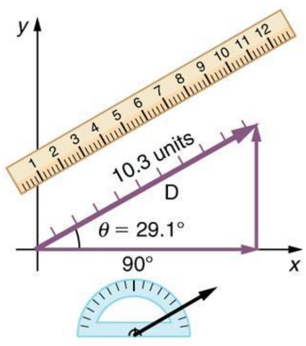


Figure 3.7:

Step 5. To get the magnitude of the resultant, *measure its length with a ruler. (Note that in most calculations, we will use the Pythagorean theorem to determine this length.)*

Step 6. To get the direction of the resultant, *measure the angle it makes with the reference frame using a protractor. (Note that in most calculations, we will use trigonometric relationships to determine this angle.)*

The graphical addition of vectors is limited in accuracy only by the precision with which the drawings can be made and the precision of the measuring tools. It is valid for any number of vectors.

Example: Adding Vectors Graphically Using the Head-to-Tail Method: A Woman Takes a Walk

Use the graphical technique for adding vectors to find the total displacement of a person who walks the following three paths (displacements) on a flat field. First, she walks 25.0 m in a direction 49.0° north of east. Then, she walks 23.0 m heading 15.0° north of east. Finally, she turns and walks 32.0 m in a direction 68.0° south of east.

Strategy

Represent each displacement vector graphically with an arrow, labeling the first **A**, the second **B**, and the third **C**, making the lengths proportional to the distance and the directions as specified relative to an east-west line. The head-to-tail method outlined above will give a way to determine the magnitude and direction of the resultant displacement, denoted **R**.

Solution

(1) Draw the three displacement vectors.

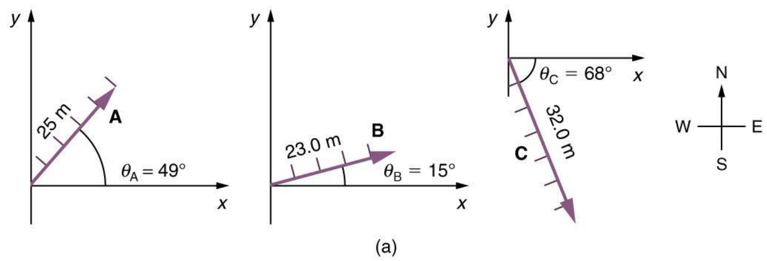


Figure 3.8:

(2) Place the vectors head to tail retaining both their initial magnitude and direction.

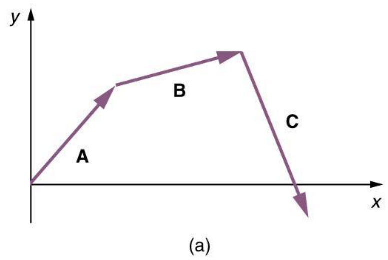


Figure 3.9:

(3) Draw the resultant vector, **R**.

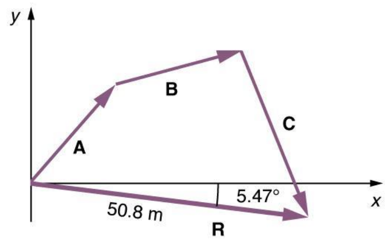


Figure 3.10:

(4) Use a ruler to measure the magnitude of **R**, and a protractor to measure the direction of **R**. While the direction of the vector can be specified in many ways, the easiest way is to measure the angle between the vector and the nearest horizontal or vertical axis. Since the resultant vector is south of the eastward pointing axis, we flip the protractor upside down and measure the angle between the eastward axis and the vector.

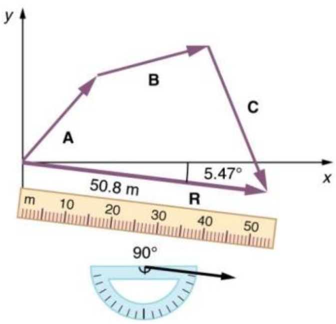


Figure 3.11:

In this case, the total displacement **R** is seen to have a magnitude of 50.8 m and to lie in a direction 5.47º south of east. By using its magnitude and direction, this vector can be expressed as R=50.8 m and θ=5.47º south of east.

Discussion

The head-to-tail graphical method of vector addition works for any number of vectors. It is also important to note that the resultant is independent of the order in which the vectors are added. This means vectors can be added in any order, but the results would be the same.

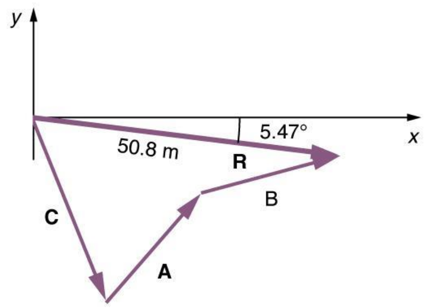


Figure 3.12:

Here, we see that when the same vectors are added in a different order, the result is the same. This characteristic is true in every case and is an important characteristic of vectors. Vector addition is commutative. Vectors can be added in any order.

$$\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A} \text{ - - - } \textbf{Equation 1}$$

(This is true for the addition of ordinary numbers as well—you get the same result whether you add **2**+**3** or **3**+**2**, for example).

3.2.3 Vector Subtraction

Vector subtraction is a straightforward extension of vector addition. To define subtraction (say we want to subtract **B** from **A** , written **A–B** , we must first define what we mean by subtraction. The *negative* of a vector **B** is defined to be **–B**; that is, graphically *the negative of any vector has the same magnitude but the opposite direction*, as shown in **Figure 3.13**. In other words, **B** has the same length as **–B**, but points in the opposite direction. Essentially, we just flip the vector so it points in the opposite direction.

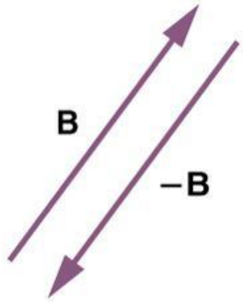


Figure 3.13: The negative of a vector is just another vector of the same magnitude but pointing in the opposite direction. So **B** is the negative of **–B**; it has the same length but opposite direction.

The *subtraction* of vector **B** from vector **A** is then simply defined to be the addition of **–B** to **A**. Note that vector subtraction is the addition of a negative vector. The order of subtraction does not affect the results.

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B})$$ - - - **Equation 2**

This is analogous to the subtraction of scalars (where, for example, $5 - 2 = 5 + (-2)$). Again, the result is independent of the order in which the subtraction is made. When vectors are subtracted graphically, the techniques outlined above are used, as the following example illustrates.

Example 3.2: Subtracting Vectors Graphically: A Woman Sailing a Boat

A woman sailing a boat at night is following directions to a dock. The instructions read to first sail 27.5 m in a direction 66.0º north of east from her current location, and then travel 30.0 m in a direction 112º north of east (or 22.0º west of north). If the woman makes a mistake and travels in the *opposite* direction for the second leg of the trip, where will she end up? Compare this location with the location of the dock.

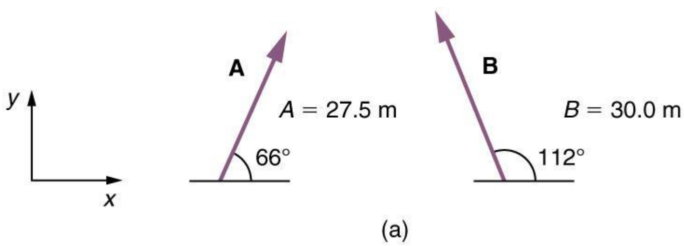


Figure 3.14:

Strategy

We can represent the first leg of the trip with a vector **A**, and the second leg of the trip with a vector **B**. The dock is located at a location **A**+**B**. If the woman mistakenly travels in the *opposite* direction for the second leg of the journey, she will travel a distance *B* (30.0 m) in the direction 180º–112º=68º south of east. We represent this as **–B**, as shown below. The vector **–B** has the same magnitude as **B** but is in the opposite direction. Thus, she will end up at a location **A**+(**–B**), or **A**–**B**.

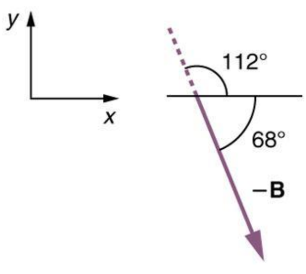


Figure 3.15:

We will perform vector addition to compare the location of the dock, **A** + B, with the location at which the woman mistakenly arrives, **A + (−B)**.

Solution

(1) To determine the location at which the woman arrives by accident, draw vectors **A** and **−B**.

(2) Place the vectors head to tail.

(3) Draw the resultant vector R.

(4) Use a ruler and protractor to measure the magnitude and direction of R.

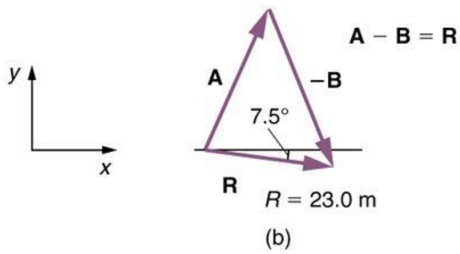


Figure 3.16:

In this case, R=23.0 m and ϑ=7.5º south of east.

(5) To determine the location of the dock, we repeat this method to add vectors **A** and **B**. We obtain the resultant vector **R'**:

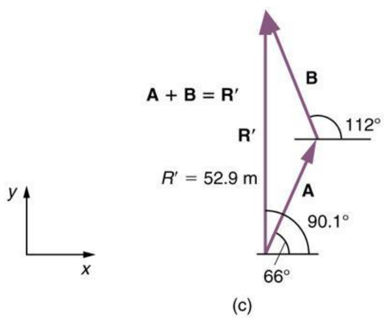


Figure 3.17:

In this case R = 52.9 m and θ=90.1º north of east.

We can see that the woman will end up a significant distance from the dock if she travels in the opposite direction for the second leg of the trip.

Discussion

Because subtraction of a vector is the same as addition of a vector with the opposite direction, the graphical method of subtracting vectors works the same as for addition.

3.3 Multiplication of Vectors and Scalars

If we decided to walk three times as far on the first leg of the trip considered in the preceding example, then we would walk $3 \times 27.5\ \text{m}$, or 82.5 m, in a direction 66.0° north of east. This is an example of multiplying a vector by a positive scalar. Notice that the magnitude changes, but the direction stays the same.

If the scalar is negative, then multiplying a vector by it changes the vector's magnitude and gives the new vector the *opposite* direction. For example, if you multiply by –2, the magnitude doubles but the direction changes. We can summarize these rules in the following way: When vector $\mathbf{A}$ is multiplied by a scalar c,

- the magnitude of the vector becomes the absolute value of *cA*,
- if *c* is positive, the direction of the vector does not change,
- if *c* is negative, the direction is reversed.

In our case, $c=3$ and $A=27.5\ \text{m}$. Vectors are multiplied by scalars in many situations. Note that division is the inverse of multiplication. For example, dividing by 2 is the same as multiplying by the value (1/2). The rules for multiplication of vectors by scalars are the same for division; simply treat the divisor as a scalar between 0 and 1.

3.4 Resolving a Vector into Components

In the examples above, we have been adding vectors to determine the resultant vector. In many cases, however, we will need to do the opposite. We will need to take a single vector and find what other vectors added together produce it. In most cases, this involves determining the perpendicular components of a single vector, for example the *x- and y*-components, or the north-south and east-west components.

For example, we may know that the total displacement of a person walking in a city is 10.3 blocks in a direction 29.0° north of east and want to find out how many blocks east and north had to be walked. This method is called *finding the components (or parts)* of the displacement in the east and north directions, and it is the inverse of the process followed to find the total displacement. It is one example of finding the components of a vector. There are many applications in physics where this is a useful thing to do.

Analytical methods of vector addition and subtraction employ geometry and simple trigonometry rather than the ruler and protractor of graphical methods. Part of the graphical technique is retained, because vectors are still represented by arrows for easy visualization. However, analytical methods are more concise, accurate, and precise than graphical methods, which are limited by the accuracy with which a drawing can be made. Analytical methods are limited only by the accuracy and precision with which physical quantities are known.

3.4.1 Resolving a Vector into Perpendicular Components

Analytical techniques and right triangles go hand-in-hand in physics because (among other things) motions along perpendicular directions are independent. We very often need to separate a vector into perpendicular components. For example, given a vector like A in Figure 3.18, we may wish to find which two perpendicular vectors, A_x and A_y, add to produce it.

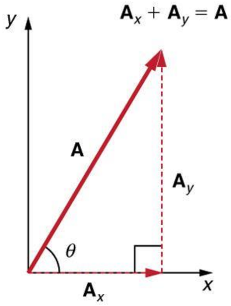


Figure 3.18: The vector A, with its tail at the origin of an *x*, *y*-coordinate system, is shown together with its *x*- and *y*-components, A_x and A_y. These vectors form a right triangle. The analytical relationships among these vectors are summarized below.

A_x and A_y are defined to be the components of A along the *x*- and *y*-axes. The three vectors A, A_x, and A_y form a right triangle:

$$A_x + \mathbf{A}_y = \mathbf{A} \text{ --- } \textbf{Equation 3}$$

Note that this relationship between vector components and the resultant vector holds only for vector quantities (which include both magnitude and direction). The relationship does not apply for the magnitudes alone. For example, if $\mathbf{A}_x = 3\text{ m}$ east, $\mathbf{A}_y = 4\text{ m}$ north, and $\mathbf{A} = 5\text{ m}$ north-east, then it is true that the vectors $A_x + \mathbf{A}_y = \mathbf{A}$. However, it is *not* true that the sum of the magnitudes of the vectors is also equal. That is,

$$3\text{ m} + 4\text{ m} \neq 5\text{ m} \text{ --- } \textbf{Equation 4}$$

Thus,

$$A_x + A_y \neq A \text{ --- } \textbf{Equation 5}$$

If the vector A is known, then its magnitude A (its length) and its angle ϑ (its direction) are known. To find A_x and A_y, its *x*- and *y*-components, we use the following relationships for a right triangle.

$$A_x = A\cos\vartheta \text{ --- } \textbf{Equation 6}$$

and

$$A_y = A\sin\vartheta$$ --- **Equation 7**

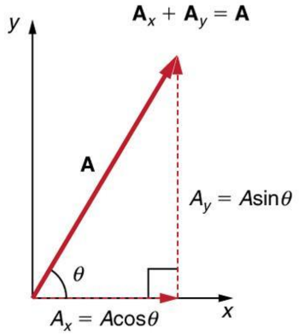


Figure 3.19: The magnitudes of the vector components A_x and A_y can be related to the resultant vector A and the angle θ with trigonometric identities. Here we see that $A_x = A\cos\theta$ and $A_y = A\sin\theta$.

Suppose, for example, that A is the vector representing the total displacement of the person walking in a city considered as given below:

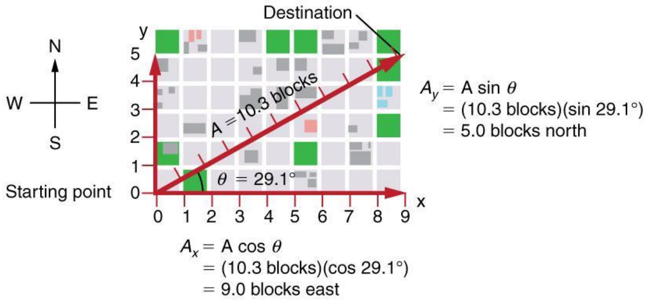


Figure 3.20: We can use the relationships $A_x = A\cos\theta$ and $A_y = A\sin\theta$ to determine the magnitude of the horizontal and vertical component vectors in this example.

Then A=10.3 blocks and θ=29.1º , so that

$$A_x = A\cos\theta = (10.3 \text{ blocks})(\cos 29.1^{\circ}) = 9.0 \text{ blocks}$$ --- **Equation 8**

$$A_y = A\sin\theta = (10.3 \text{ blocks})(\sin 29.1^{\circ}) = 5.0 \text{ blocks}$$ --- **Equation 9**

3.4.1.1 Calculating a Resultant Vector

If the perpendicular components A*x* and A*y* of a vector A are known, then A can also be found analytically. To find the magnitude *A* and direction θ of a vector from its perpendicular components A*x* and A*y*, relative to the *x*-axis, we use the following relationships:

$$A = \sqrt{Ax\char`^2 + Ay\char`^2} \quad \text{--- } \textbf{Equation 10}$$

$$\theta = \tan^{-1}(A_y/A_x) \text{ --- } \textbf{Equation 11}$$

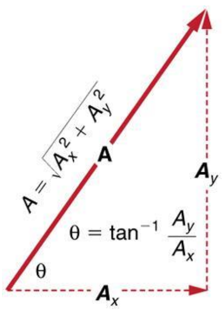


Figure 3.21: The magnitude and direction of the resultant vector can be determined once the horizontal and vertical components *Ax* and *Ay* have been determined.

3.5 Determining Vectors and Vector Components with Analytical Methods

Equations $A_x = A\cos\theta$ and $A_y = A\sin\theta$ are used to find the perpendicular components of a vector—that is, to go from A and θ to A_x and A_y. Equations $A = \sqrt{Ax\char`^2 + Ay\char`^2}$ and $\theta = \tan^{-1}(A_y/A_x)$ are used to find a vector from its perpendicular components—that is, to go from A_x and A_y to A and θ. Both processes are crucial to analytical methods of vector addition and subtraction.

3.5.1 Adding Vectors Using Analytical Methods

To see how to add vectors using perpendicular components, consider Figure 3.22, in which the vectors A and B are added to produce the resultant R.

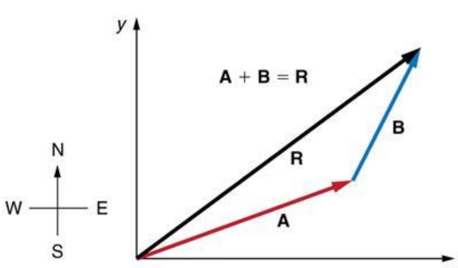


Figure 3.22: Vectors A and B are two legs of a walk, and R is the resultant or total displacement. You can use analytical methods to determine the magnitude and direction of R.

If A and B represent two legs of a walk (two displacements), then R is the total displacement. The person taking the walk ends up at the tip of R. There are many ways to arrive at the same point. In particular, the person could have walked first in the *x*-direction and then in the *y*-direction. Those paths are the *x*- and *y*-components of the resultant, R*x* and R*y*. If we know **R***x* and R*y*, we can find *R* and θ using the equations $A = \sqrt{Ax\char`^2 + Ay\char`^2}$ and $\theta = \tan{-1}(A_y/A_x)$. When you use the analytical method of vector addition, you can determine the components or the magnitude and direction of a vector.

Step 1. *Identify the x- and y-axes that will be used in the problem. Then, find the components of each vector to be added along the chosen perpendicular axes.* Use the equations $A_x = A\cos\theta$ and $A_y = A\sin\theta$ to find the components. In Figure 3.23, these components are *Ax*, *Ay*, *Bx*, and *By*. The angles that vectors A and B make with the *x*-axis are θ_A and θ_B, respectively.

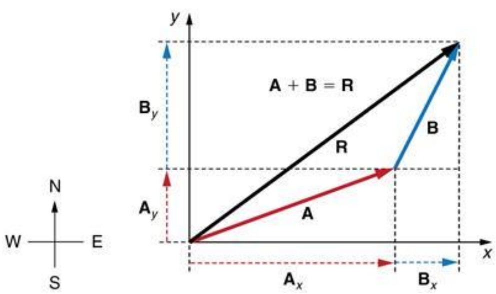


Figure 3.23: To add vectors A and B, first determine the horizontal and vertical components of each vector. These are the dotted vectors A_x, A_y, B_x and B_y shown in the image.

Step 2. *Find the components of the resultant along each axis by adding the components of the individual vectors along that axis.* That is, as shown in Figure 3.24,

$$R_x = A_x + B_x$$ --- **Equation 12**

and

$$R_y = A_y + B_y$$ --- **Equation 13**

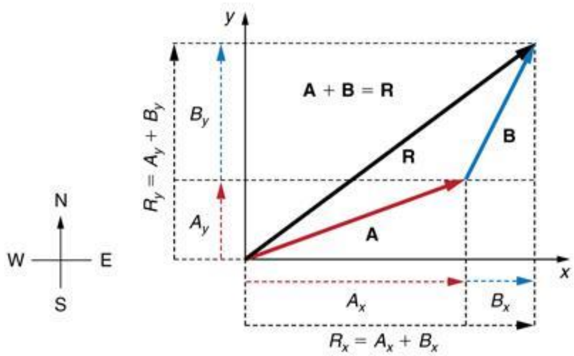


Figure 3.24: The magnitude of the vectors A*x* and B*x* add to give the magnitude *Rx* of the resultant vector in the horizontal direction. Similarly, the magnitudes of the vectors A*y* and B*y* add to give the magnitude *Ry* of the resultant vector in the vertical direction.

Components along the same axis, say the *x*-axis, are vectors along the same line and, thus, can be added to one another like ordinary numbers. The same is true for components along the *y*-axis. (For example, a 9-block eastward walk could be taken in two legs, the first 3 blocks east and the second 6 blocks east, for a total of 9, because they are along the same direction.) So resolving vectors into components along common axes makes it easier to add them. Now that the components of R are known, its magnitude and direction can be found.

Step 3. *To get the magnitude R of the resultant, use the Pythagorean theorem:*

$$R = \sqrt{Rx\text{^}2 + Ry\text{^}2} \text{ --- } \textbf{Equation 14}$$

Step 4. *To get the direction of the resultant relative to the x-axis:*

$$\vartheta = \tan^{-1}(R_y/R_x) \text{ --- } \textbf{Equation 15}$$

The following example illustrates this technique for adding vectors using perpendicular components.

An Example on Adding Vectors Using Analytical Methods

Add the vector A to the vector B shown in Figure 3.25, using perpendicular components along the *x*- and *y*-axes. The *x*- and *y*-axes are along the east–west and north–south directions, respectively. Vector A represents the first leg of a walk in which a person walks 53.0 m in a direction 20.0º north of east. Vector B represents the second leg, a displacement of 34.0 m in a direction 63.0º north of east.

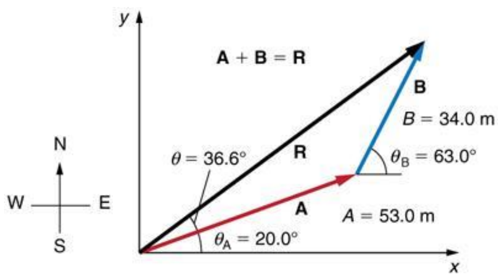


Figure 3.24: Vector A has magnitude 53.0 m and direction 20.0º north of the *x*-axis. Vector B has magnitude 34.0 m and direction 63.0º north of the *x*-axis. You can use analytical methods to determine the magnitude and direction of R.

Strategy

The components of A and B along the *x*- and *y*-axes represent walking due east and due north to get to the same ending point. Once found, they are combined to produce the resultant.

Solution

Following the method outlined above, we first find the components of A and B along the *x*- and *y*-axes. Note that A=53.0 m, $\theta_A = 20.0º$, B = 34.0 m, and $\theta_B = 63.0º$. We find the *x*-components by using $Ax=A\cos\theta$, which gives

$$A_x = A\cos\theta_A = (53.0 \text{ m})(\cos 20.0º)(53.0 \text{ m})(0.940) = 49.8 \text{ m}$$

and

$$B_x = B\cos\theta_B = (34.0 \text{ m})(\cos 63.0º)(34.0 \text{ m})(0.454) = 15.4 \text{ m}.$$

Similarly, the *y*-components are found using $A_y = A\sin\theta_A$:

$$A_y = A\sin\theta_A = (53.0 \text{ m})(\sin 20.0º)(53.0 \text{ m})(0.342) = 18.1 \text{ m}$$

and

$$B_y = B\sin\theta_B = (34.0 \text{ m})(\sin 63.0º)(34.0 \text{ m})(0.891) = 30.3 \text{ m}.$$

The *x*- and *y*-components of the resultant are thus,

$$R_x = A_x + B_x = 49.8 \text{ m} + 15.4 \text{ m} = 65.2 \text{ m}$$

and

$$R_y = A_y + B_y = 18.1\text{ m} + 30.3\text{ m} = 48.4\text{ m}.$$

Now we can find the magnitude of the resultant by using the Pythagorean theorem:

$$R = \sqrt{(Rx^2 + Ry^2)} = \sqrt{(65.2)^2+(48.4)^2}\,\text{m}$$

so that

$$R = 81.2\text{ m}.$$

Finally, we find the direction of the resultant:

$$\Theta = \tan^{-1}(Ry/Rx) = +\tan^{-1}(48.4/65.2).$$

Thus,

$$\Theta = \tan^{-1}(0.742) = 36.6^{\circ}.$$

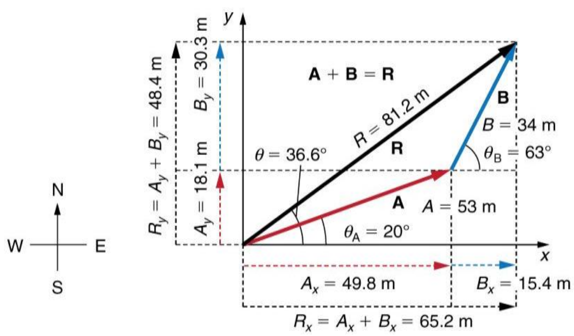


Figure 3.25: Using analytical methods, we see that the magnitude of R is 81.2 m and its direction is 36.6º north of east.

3.6 Projectile Motion

Projectile motion is a form of motion where an object moves in a bilaterally symmetrical, parabolic path. The path that the object follows is called its trajectory. Projectile motion only occurs when there is one force applied at the beginning on the trajectory, after which the only interference is from gravity.

The motion of falling objects, as covered under one dimensional kinematics is a simple one-dimensional type of projectile motion in which there is no horizontal movement. In this section, we consider two-dimensional projectile motion, such as that of a football or other object for which air resistance *is negligible*.

The most important fact to remember here is that *motions along perpendicular axes are independent* and thus can be analyzed separately. This fact was discussed under kinematics in two dimensions, where vertical and horizontal motions were seen to be independent. The key to analyzing two-dimensional projectile motion is to break it into two motions, one along the horizontal axis and the other along the vertical. (This choice of axes is the most sensible, because acceleration due to gravity is vertical—thus, there will be no acceleration along the horizontal axis when air resistance is negligible.) As is customary, we call the horizontal axis the *x*-axis and the vertical axis the *y*-axis. Figure 3.26 illustrates the notation for displacement, where s is defined to be the total displacement and x and y are its components along the horizontal and vertical axes, respectively. The magnitudes of these vectors are *s*, *x*, and *y*. (Note that in the last section we used the notation A to represent a vector with components A_x and A_y. If we continued this format, we would call displacement s with components s_x and s_y. However, to simplify the notation, we will simply represent the component vectors as x and y.)

Of course, to describe motion we must deal with velocity and acceleration, as well as with displacement. We must find their components along the *x*- and *y*-axes, too. We will assume all forces except gravity (such as air resistance and friction, for example) are negligible. The components of acceleration are then very simple: $a_y = -g = -9.80 \text{ m/s}^2$. (Note that this definition assumes that the upwards direction is defined as the positive direction. If you arrange the coordinate system instead such that the downwards direction is positive, then acceleration due to gravity takes a positive value.) Because gravity is vertical, $a_x = 0$. Both accelerations are constant, so the kinematic equations can be used.

The following are the Kinematic Equations (at constant *a*) already discussed:

$$x = x_0 + \bar{v}t$$

$$v_{average} = (v_0 + v) / 2$$

$$v = v_0 + at$$

$$v^2 = v_0^2 + 2a(x - x_0)$$

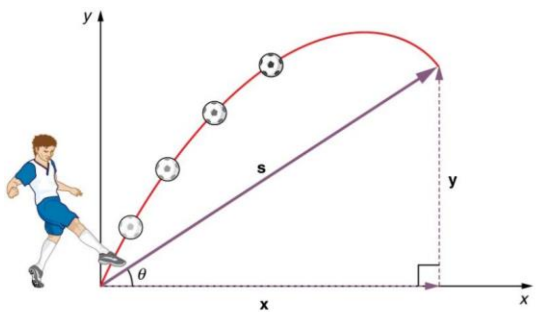


Figure 3.26: The total displacement s of a soccer ball at a point along its path. The vector s has components x and y along the horizontal and vertical axes. Its magnitude is *s*, and it makes an angle θ with the horizontal.

Given these assumptions, the following steps are then used to analyze projectile motion:

Step 1. *Resolve or break the motion into horizontal and vertical components along the x- and y-axes.* These axes are perpendicular, so $A_x = A\cos\theta$ and $A_y = A\sin\theta$ are used. The magnitude of the components of displacement s along these axes are *x* and *y*. The magnitudes of the components of the velocity v are $v_x = v\cos\theta$ and $v_y = v\sin\theta$, where *v* is the magnitude of the velocity and θ is its direction, as shown in Figure 3.27. Initial values are denoted with a subscript 0, as usual.

Step 2. *Treat the motion as two independent one-dimensional motions, one horizontal and the other vertical.* The kinematic equations for horizontal and vertical motion take the following forms:

Horizontal Motion ($a_x = 0$)

$$x = x_0 + v_x t$$

$$v_x = v_{0x} = v_x = \text{velocity is constant.}$$

Vertical Motion (assuming positive is up)

$$a_y = -g = -9.80\text{m/s}^2$$

$$y = y_0 + 1/2\,(v_{0y}+v_y)t$$

$$v_y = v_{0y} - gt$$

$$y = y_0 + v_{0y}t - 1/2gt^2$$

$$v^2{}_y = v^2{}_{0y} - 2g\,(y - y_0)$$

Step 3. *Solve for the unknowns in the two separate motions—one horizontal and one vertical.* Note that the only common variable between the motions is time *t*. The problem-solving procedures here are the same as for one-dimensional kinematics and are illustrated in the solved examples below.

Step 4. *Recombine the two motions to find the total displacement* **s** *and velocity* **v**. Because the *x* - and *y* -motions are perpendicular, we determine these vectors by using the techniques outlined in the Vector Addition and Subtraction methods and employing,

$A = \sqrt{Ax^2 + Ay^2}$ and $\theta = \tan^{-1}(A_y/A_x)$ in the following form, where θ is the direction of the displacement s and θ_v is the direction of the velocity v:

Total displacement and velocity

$$s = \sqrt{x^2 + y^2}$$ --- **Equation 16**

$$\theta = \tan^{-1}(y/x)$$ --- **Equation 17**

$$v = \sqrt{v_x^2 + v_y^2}$$ --- **Equation 18**

$$\theta_v = \tan^{-1}(v_y/v_x)$$ --- **Equation 19**

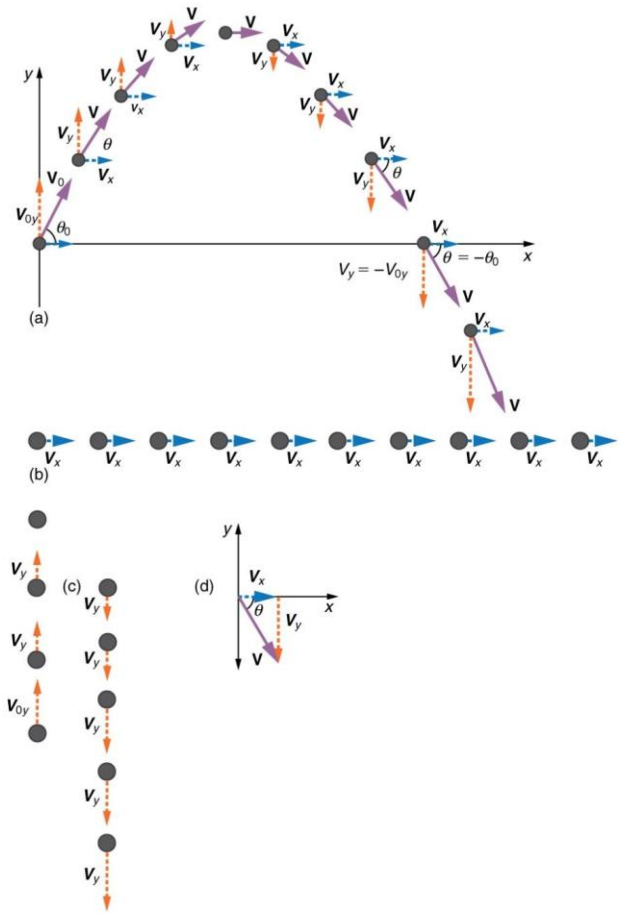


Figure 3.27: (a) We analyze two-dimensional projectile motion by breaking it into two independent one-dimensional motions along the vertical and horizontal axes. (b) The horizontal motion is simple, because $a_x=0$ and v_x is thus constant. (c) The velocity in the vertical direction begins to decrease as the object rises; at its highest point, the vertical velocity is zero. As the object falls towards the Earth again, the vertical velocity increases again in magnitude but points in the opposite direction to the initial vertical velocity. (d) The x - and y -motions are recombined to give the total velocity at any given point on the trajectory.

Defining a Coordinate System

It is important to set up a coordinate system when analyzing projectile motion. One part of defining the coordinate system is to define an origin for the x and y positions. Often, it is convenient to choose the initial position of the object as the origin such that $x_0 = 0$ and $y_0 = 0$. It is also important to define the positive and negative directions in the x and y directions. Typically, we define the positive vertical direction as upwards, and the positive horizontal direction is usually the direction of the object's motion. When this is the case, the vertical acceleration, g, takes a negative value (since it is directed downwards towards the Earth). However, it is occasionally useful to define the coordinates differently. For example, if you are analyzing the motion of a ball thrown downwards from the top of a cliff, it may make sense to define the positive direction downwards since the motion of the ball is solely in the downwards direction. If this is the case, g takes a positive value.

3.7 Relative Velocity

If a person rows a boat across a rapidly flowing river and tries to head directly for the other shore, the boat instead moves *diagonally* relative to the shore, as in Figure 3.28. The boat does not move in the direction in which it is pointed. The reason, of course, is that the river carries the boat downstream. Similarly, if a small airplane flies overhead in a strong crosswind, you can sometimes see that the plane is not moving in the direction in which it is pointed, as illustrated in Figure 3.41. The plane is moving straight ahead relative to the air, but the movement of the air mass relative to the ground carries it sideways.

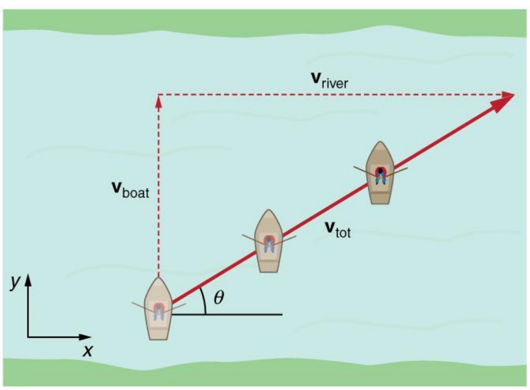


Figure 3.28: A boat trying to head straight across a river will actually move diagonally relative to the shore as shown. Its total velocity (solid arrow) relative to the shore is the sum of its velocity relative to the river plus the velocity of the river relative to the shore.

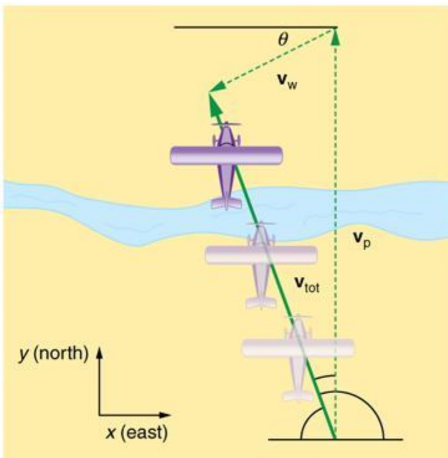


Figure 3.29: An airplane heading straight north is instead carried to the west and slowed down by wind. The plane does not move relative to the ground in the direction it points; rather, it moves in the direction of its total velocity (solid arrow).

In each of these situations, an object has a velocity relative to a medium (such as a river) and that medium has a velocity relative to an observer on solid ground. The velocity of the object *relative to the observer* is the sum of these velocity vectors, as indicated in Figures 3.28 and 3.29. These situations are only two of many in which it is useful to add velocities.

Velocity is a vector (it has both magnitude and direction); the rules of vector addition discussed in Vector Addition and Subtraction: Graphical Methods and Vector Addition and Subtraction: Analytical Methods apply to the addition of velocities, just as they do for any other vectors. In one-dimensional motion, the addition of velocities is simple—they add like ordinary numbers. For example, if a field hockey player is moving at 5 m/s straight toward the goal and drives the ball in the same direction with a velocity of 30 m/s relative to her body, then the velocity of the ball is 35 m/s relative to the stationary, profusely sweating goalkeeper standing in front of the goal.

In two-dimensional motion, either graphical or analytical techniques can be used to add velocities. We will concentrate on analytical techniques. The following equations give the relationships between the magnitude and direction of velocity (v and ϑ) and its components (v_x and v_y) along the *x*- and *y*-axes of an appropriately chosen coordinate system:

$$v_x = v\cos\vartheta$$

$$v_y = v\sin\vartheta$$

$$v = \sqrt{vx^2 + vy^2}$$

$$\vartheta = \tan^{-1}(v_y/v_x)$$

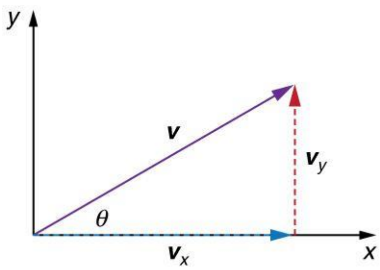


Figure 3.30: The velocity, *v*, of an object traveling at an angle θ to the horizontal axis is the sum of component vectors $\mathbf{v}_x$ and $\mathbf{v}_y$.

These equations are valid for any vectors and are adapted specifically for velocity. The first two equations are used to find the components of a velocity when its magnitude and direction are known. The last two are used to find the magnitude and direction of velocity when its components are known.

Example on Adding Velocities: A Boat on a River

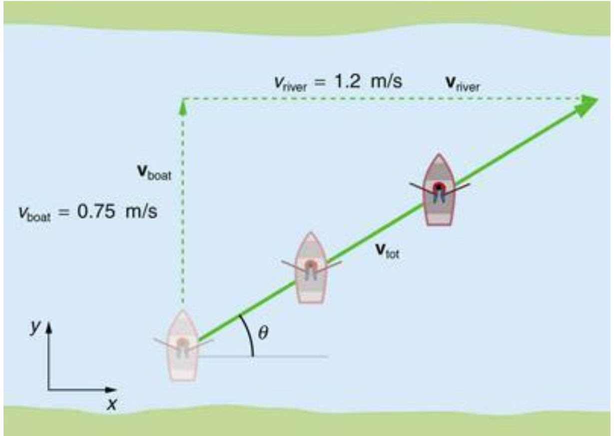


Figure 3.31: A boat attempts to travel straight across a river at a speed 0.75 m/s. The current in the river, however, flows at a speed of 1.20 m/s to the right.

Refer to Figure 3.31, which shows a boat trying to go straight across the river. Let us calculate the magnitude and direction of the boat's velocity relative to an observer on the shore, **v**tot. The velocity of the boat, **v**boat, is 0.75 m/s in the *y*-direction relative to the river and the velocity of the river, **v**river, is 1.20 m/s to the right.

Strategy

We start by choosing a coordinate system with its *x*-axis parallel to the velocity of the river, as shown in Figure 3.31. Because the boat is directed straight toward the other shore, its velocity relative to the water is parallel to the *y*-axis and perpendicular to the velocity of the river.

Thus, we can add the two velocities by using the equations $v_{\text{tot}} = \sqrt{v^2{}_x + v^2{}_y}$ and $\theta = \tan^{-1}(v_y/v_x)$ directly.

Solution

The magnitude of the total velocity is

$$v_{\text{tot}} = \sqrt{v^2{}_x + v^2{}_y}$$

where

$$v_x = v_{\text{river}} = 1.20 \text{ m/s}$$

and

$$v_y = v_{\text{boat}} = 0.750 \text{ m/s}.$$

Thus,

$$v_{\text{tot}} = \sqrt{(1.20 \text{ m/s})\hat{}2 + (0.750 \text{ m/s})\hat{}2}$$

yielding

$$v_{\text{tot}} = 1.42 \text{ m/s}.$$

The direction of the total velocity θ is given by:

$$\theta = \tan^{-1}(v_y/v_x) = \tan^{-1}(0.750/1.20).$$

This equation gives,

$$\theta = 32.0^{\circ}.$$

Discussion

Both the magnitude *v* and the direction θ of the total velocity are consistent with Figure 3.31. Note that because the velocity of the river is large compared with the velocity of the boat, it is swept rapidly downstream. This result is evidenced by the small angle (only 32.0º) the total velocity has relative to the riverbank.

Example on Calculating Velocity: Wind Velocity Causes an Airplane to Drift

Calculate the wind velocity for the situation shown in Figure 3.32. The plane is known to be moving at 45.0 m/s due north relative to the air mass, while its velocity relative to the ground (its total velocity) is 38.0 m/s in a direction 20.0º west of north.

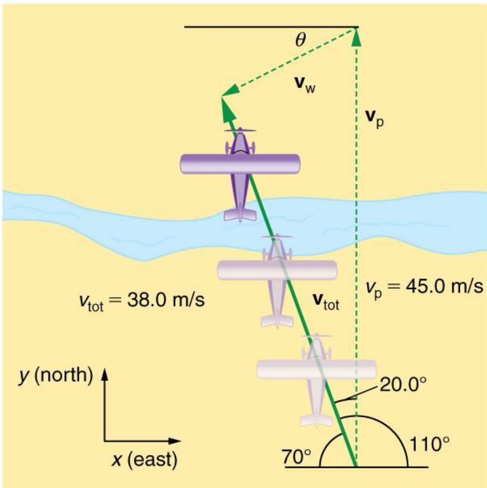


Figure 3.32: An airplane is known to be heading north at 45.0 m/s, though its velocity relative to the ground is 38.0 m/s at an angle west of north. What is the speed and direction of the wind?

Strategy

In this problem, somewhat different from the previous example, we know the total velocity $\mathbf{v}_{tot}$ and that it is the sum of two other velocities, $\mathbf{v}_w$ (the wind) and $\mathbf{v}_p$ (the plane relative to the air mass). The quantity $\mathbf{v}_p$ is known, and we are asked to find $\mathbf{v}_w$. None of the velocities are perpendicular, but it is possible to find their components along a common set of perpendicular axes. If we can find the components of **v**w, then we can combine them to solve for its magnitude and direction. As shown in Figure 3.32, we choose a coordinate system with its *x*-axis due east and its *y*-axis due north (parallel to $\mathbf{v}_p$)

Solution

Because $\mathbf{v}_{tot}$ is the vector sum of the $\mathbf{v}_w$ and $\mathbf{v}_p$, its *x*- and *y*-components are the sums of the *x*- and *y*-components of the wind and plane velocities. Note that the plane only has vertical component of velocity so $v_{px}=0$ and $v_{py}=v_p$. That is,

$$v_{totx} = v_{wx}$$

and

$$v_{toty}=v_{wy}+v_p$$

We can use the first of these two equations to find v_{wx}:

$$v_{wx}= v_{totx}= v_{tot}\cos 110^o$$

Because $v_{tot} = 38.0$ m/s and $\cos 110^o = -0.342$, we have

$$v_{wx} = (38.0 \text{ m/s})(–0.342) = –13 \text{ m/s}$$

The minus sign indicates motion west which is consistent with the diagram.

Now, to find v_{wy} we note that,

$$v_{tot y} = v_{wy} + v_p$$

Here $v_{tot\,y} = v_{tot} \sin 110^o$; thus,

$$v_{wy} = (38.0 \text{ m/s})(0.940) − 45.0 \text{ m/s} = −9.29 \text{ m/s}.$$

This minus sign indicates motion south which is consistent with the diagram.

Now that the perpendicular components of the wind velocity v_{wx} and v_{wy} are known, we can find the magnitude and direction of $\mathbf{v}_w$. First, the magnitude is,

$$v_w = \sqrt{v^2_{wx} + v^2_{wy}}$$

$$\sqrt{(−13.0 \text{ m/s})^2 + (−9.29 \text{ m/s})^2}$$

so that

$$v_w = 16.0 \text{ m/s}.$$

The direction is:

$$\theta = \tan^{-1}(v_{wy}/v_{wx}) = \tan^{-1}(−9.29/−13.0)$$

giving

$$\vartheta = 35.6^o.$$

Relative Velocities and Classical Relativity

When adding velocities, we have been careful to specify that the *velocity is relative to some reference frame*. These velocities are called relative velocities. For example, the velocity of an airplane relative to an air mass is different from its velocity relative to the ground. Both are quite different from the velocity of an airplane relative to its passengers (which should be close to zero). Relative velocities are one aspect of relativity, which is defined to be the study of how different observers moving relative to each other measure the same phenomenon.

Let us consider an example of what two different observers see in a situation analyzed long ago by Galileo. Suppose a sailor at the top of a mast on a moving ship drops their binoculars. Where will it hit the deck? Will it hit at the base of the mast, or will it hit behind the mast because the ship is moving forward? The answer is that if air resistance is negligible, the binoculars will hit at the base of the mast at a point directly below its point of release. Now let us consider what two different observers see when the binoculars drop. One observer is on the ship and the other on shore. The binoculars have no horizontal velocity relative to the observer on the ship, and so he sees them fall straight down the mast. (See Figure 3.33) To the observer on shore, the binoculars and the ship have the *same* horizontal velocity, so both move the same distance forward while the binoculars are falling. This observer sees the curved path shown in Figure 3.33. Although the paths look different to the different observers, each sees the same result—the binoculars hit at the base of the mast and not behind it. To get the correct description, it is crucial to correctly specify the velocities relative to the observer.

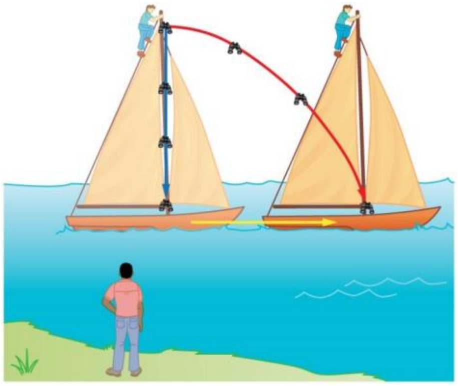


Figure 3.33: Classical relativity. The same motion as viewed by two different observers. An observer on the moving ship sees the binoculars dropped from the top of its mast fall straight down. An observer on shore sees the binoculars take the curved path, moving forward with the ship. Both observers see the binoculars strike the deck at the base of the mast. The initial horizontal velocity is different relative to the two observers. (The ship is shown moving rather fast to emphasize the effect.)

Example of Calculating Relative Velocity: An Airline Passenger Drops a Coin

An airline passenger drops a coin while the plane is moving at 260 m/s. What is the velocity of the coin when it strikes the floor 1.50 m below its point of release: (a) Measured relative to the plane? (b) Measured relative to the Earth?

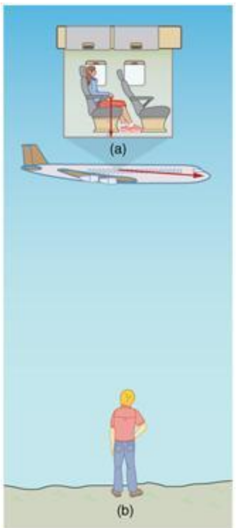


Figure 3.34: The motion of a coin dropped inside an airplane as viewed by two different observers. (a) An observer in the plane sees the coin fall straight down. (b) An observer on the ground sees the coin move almost horizontally.

Strategy

Both problems can be solved with the techniques for falling objects and projectiles. In part (a), the initial velocity of the coin is zero relative to the plane, so the motion is that of a falling object (one-dimensional). In part (b), the initial velocity is 260 m/s horizontal relative to the Earth and gravity is vertical, so this motion is a projectile motion. In both parts, it is best to use a coordinate system with vertical and horizontal axes.

Solution for (a)

Using the given information, we note that the initial velocity and position are zero, and the final position is 1.50 m. The final velocity can be found using the equation:

$$v_y^2 = v_{0,y}^2 - 2g(y - y_0).$$

Substituting known values into the equation, we get

$$v_y^2 = 0^2 - 2(9.80\text{m/s}^2)(-1.50\text{m} - 0\text{ m}) = 29.4\text{m}^2/\text{s}^2$$

yielding

$$v_y = -5.42\text{ m/s}.$$

We know that the square root of 29.4 has two roots: 5.42 and -5.42. We choose the negative root because we know that the velocity is directed downwards, and we have defined the positive direction to be upwards. There is no initial horizontal velocity relative to the plane and no horizontal acceleration, and so the motion is straight down relative to the plane.

Solution for (b)

Because the initial vertical velocity is zero relative to the ground and vertical motion is independent of horizontal motion, the final vertical velocity for the coin relative to the ground is $v_y = -5.42 \text{m/s}$, the same as found in part (a). In contrast to part (a), there now is a horizontal component of the velocity. However, since there is no horizontal acceleration, the initial and final horizontal velocities are the same and $v_x = 260$ m/s. The *x*- and *y*-components of velocity can be combined to find the magnitude of the final velocity:

$$v = \sqrt{v_x^2 + v_y^2}$$

Thus,

$$v = \sqrt{(260 \text{ m/s})^2 + (-5.42 \text{ m/s})^2}$$

yielding

$$v = 260.06 \text{ m/s}.$$

The direction is given by:

$$\theta = \tan^{-1}(v_y/v_x) = \tan^{-1}(-5.42/260)$$

so that,

$$\theta = \tan^{-1}(-0.0208) = -1.19^{\circ}$$

Chapter 4

4.0 Objectives

At the end of this lesson, students should be able to,

- Define mass and inertia.
- Apply Newton's first law of motion.
- Identify net force, external force, and system.
- Apply Newton’s second law to solve problems.
- Apply Newton's third law to define systems and solve problems of motion.

4.1 Introduction

This chapter covers Newton’s three laws and their applicability in solving problems related to motion, force, friction, etc.

The study of motion is kinematics, but kinematics only *describes* the way objects move—their velocity and their acceleration. Dynamics considers the forces that affect the motion of moving objects and systems. Newton’s laws of motion are the foundation of dynamics. These laws provide an example of the breadth and simplicity of principles under which nature functions. They are also universal laws in that they apply to similar situations on Earth as well as in space.

Dynamics is the study of the forces that cause objects and systems to move. To understand this, we need a working definition of force. Our intuitive definition of force—that is, a push or a pull—is a good place to start. We know that a push or pull has both magnitude and direction (therefore, it is a vector quantity) and can vary considerably in each regard. If two people push in different directions on a third person, as illustrated in Figure 4.1, we might expect the total force to be in the direction shown. Since force is a vector, it adds just like other vectors, as illustrated in Figure (a) for two ice skaters. Forces, like other vectors, are represented by arrows and can be added using the familiar head-to-tail method or by trigonometric methods. This is similar to two dimensional kinematics.

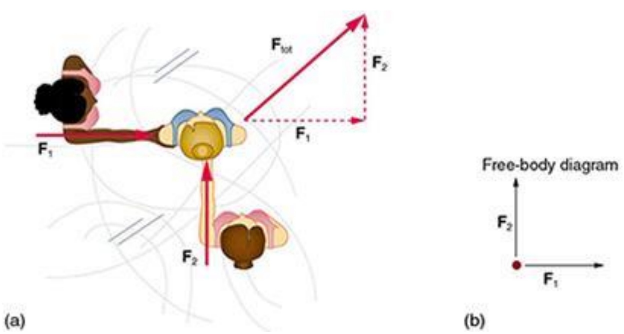


Figure 4.1: Part (a) shows an overhead view of two ice skaters pushing on a third. Forces are vectors and add like other vectors, so the total force on the third skater is in the direction shown. In part (b), we see a free-body diagram representing the forces acting on the third skater.

Figure 4.1: (b) is our first example of a free-body diagram, which is a technique used to illustrate all the external forces acting on a body. The body is represented by a single isolated point (or free body), and only those forces acting *on* the body from the outside (external forces) are shown. (These forces are the only ones shown, because only external forces acting on the body affect its motion. We can ignore any internal forces within the body.) Free-body diagrams are very useful in analyzing forces acting on a system and are employed extensively in the study and application of Newton’s laws of motion.

A more quantitative definition of force can be based on some standard force, just as distance is measured in units relative to a standard distance. One possibility is to stretch a spring a certain fixed distance, as illustrated in Figure 4.2, and use the force it exerts to pull itself back to its relaxed shape—called a *restoring force*—as a standard. The magnitude of all other forces can be stated as multiples of this standard unit of force.

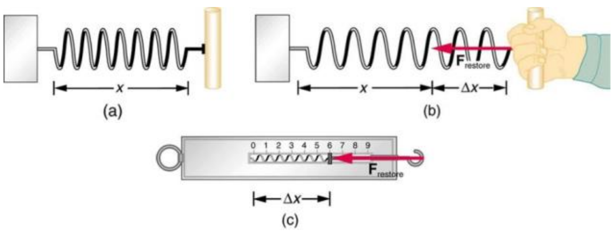


Figure 4.2: The force exerted by a stretched spring can be used as a standard unit of force. (a) This spring has a length *x* when undistorted. (b) When stretched a distance Δ*x*, the spring exerts a restoring force, $\mathbf{F}_{restore}$, which is reproducible. (c) A spring scale is one device that uses a spring to measure force. The force $\mathbf{F}_{restore}$ is exerted on whatever is attached to the hook. Here $\mathbf{F}_{restore}$ has a magnitude of 6 units in the force standard being employed.

4.2 Newton's First Law of Motion

A body at rest remains at rest, or, if in motion, remains in motion at a constant velocity unless acted on by a net external force.

Newton's first law of motion states that there must be a *cause* (which is a net external force) *for there to be any change in velocity (either a change in magnitude or direction).* An object sliding across a table or floor slows down due to the net force of friction acting on the object. If it is a frictionless surface, we can imagine the object sliding in a straight line indefinitely. Friction is thus the *cause* of the slowing (consistent with Newton's first law). Consider an air hockey table. When the air is turned off, the puck slides only a short distance before friction slows it to a stop. However, when the air is turned on, it creates a nearly frictionless surface, and the puck glides long distances without slowing down. Additionally, if we know enough about the friction, we can accurately predict how quickly the object will slow down. Friction is an external force.

Newton's first law is completely general and can be applied to anything from an object sliding on a table to a satellite in orbit to blood pumped from the heart. Experiments have thoroughly verified that any change in velocity (speed or direction) must be caused by an external force. The idea of *generally applicable or universal laws* is important not only here—it is a basic feature of all laws of physics.

Mass

The property of a body to remain at rest or to remain in motion with constant velocity is called inertia. Newton's first law is often called the law of inertia. As we know from experience, some objects have more inertia than others. It is obviously more difficult to change the motion of a large boulder than that of a basketball, for example. The inertia of an object is measured by its mass. Roughly speaking, mass is a measure of the amount of "stuff" (or matter) in something. The quantity or amount of matter in an object is determined by the numbers of atoms and molecules of various types it contains. Unlike weight, mass does not vary with location. The mass of an object is the same on Earth, in orbit, or on the surface of the Moon. In practice, it is very difficult to count and identify all of the atoms and molecules in an object, so masses are not often determined in this manner. Operationally, the masses of objects are determined by comparison with the standard kilogram.

Newton's second law of motion is closely related to Newton's first law of motion. It mathematically states the cause and effect relationship between force and changes in motion. Newton's second law of motion is more quantitative and is used extensively to calculate what happens in situations involving a force. Before we can write down Newton's second law as a simple equation giving the exact relationship of force, mass, and acceleration, we need to sharpen some ideas that have already been mentioned.

First, what do we mean by a change in motion? The answer is that a change in motion is equivalent to a change in velocity. A change in velocity means, by definition, that there is an acceleration (or change of direction). Newton's first law says that a net external force causes a change in motion; thus, we see that a *net external force causes acceleration*.

Another question immediately arises. What do we mean by an external force? An intuitive notion of external is correct—an external force acts from outside the system (object or collection of objects) of interest. For example, in Figure 4.3 (a) the system of interest is the wagon plus the child in it. The two forces exerted by the other children are external forces. An internal force acts between elements of the system. Again looking at Figure 4.3 (a), the force the child in the wagon exerts to hang onto the wagon is an internal force between elements of the system of interest. Only external forces affect the motion of a system, according to Newton's first law. (The internal forces actually cancel) *You must define the boundaries of the system before you can determine which forces are external*. Sometimes the system is obvious, whereas other times identifying the boundaries of a system is more subtle. The concept of a system is fundamental to many areas of physics, as is the correct application of Newton's laws. This concept will be revisited many times on our journey through physics.

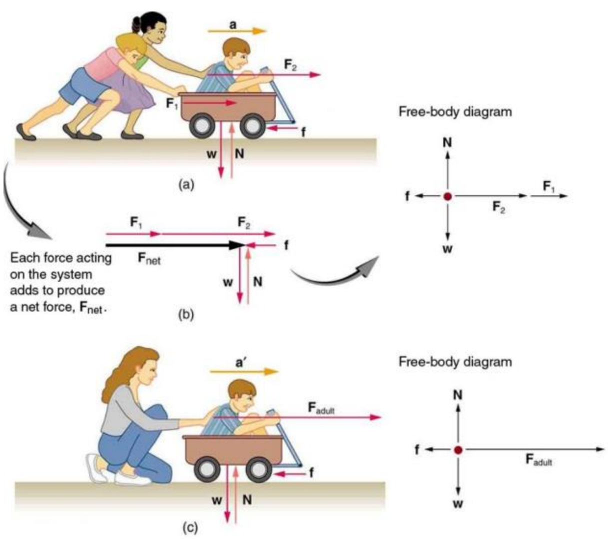


Figure **4.3:** Different forces exerted on the same mass produce different accelerations. (a) Two children push a wagon with a child in it. Arrows representing all external forces are shown. The system of interest is the wagon and its rider. The weight **w** of the system and the support of the ground **N** are also shown for completeness and are assumed to cancel. The vector **f** represents the friction acting on the wagon, and it acts to the left, opposing the motion of the wagon. (b) All of the external forces acting on the system add together to produce a net force, $\mathbf{F}_{net}$. The free-body diagram shows all of the forces acting on the system of interest. The dot represents the center of mass of the system. Each force vector extends from this dot. Because there are two forces acting to the right, we draw the vectors collinearly. (c) A larger net external force produces a larger acceleration ($\mathbf{a'} > \mathbf{a}$) when an adult pushes the child.

Now, it seems reasonable that acceleration should be directly proportional to and in the same direction as the net (total) external force acting on a system. This assumption has been verified experimentally and is illustrated in Figure 4.3. In part (a), a smaller force causes a smaller acceleration than the larger force illustrated in part (c). For completeness, the vertical forces are also shown; they are assumed to cancel since there is no acceleration in the vertical direction. The vertical forces are the weight **w** and the support of the ground **N**, and the horizontal force **f** represents the force of friction. These will be discussed in more detail in later sections. For now, we will define friction as a force that opposes the motion past each other of objects that are touching. Figure 4.3(b) shows how vectors representing the external forces add together to produce a net force, $\mathbf{F}_{net}$.

To obtain an equation for Newton's second law, we first write the relationship of acceleration and net external force as the proportionality,

$$\mathbf{a} \propto \mathbf{F}_{net} \text{ --- } \textbf{Equation 1}$$

where the symbol $\propto$ means "proportional to," and $\mathbf{F}_{net}$ is the net external force. (The net external force is the vector sum of all external forces and can be determined graphically, using the head-to-tail method, or analytically, using components. The techniques are the same as for the addition of other vectors, and are covered under two-dimensional kinematics). This proportionality states what we have said in words—*acceleration is directly proportional to the net external force.* Once the system of interest is chosen, it is important to identify the external forces and ignore the internal ones. It is a tremendous simplification not to have to consider the numerous internal forces acting between objects within the system, such as muscular forces within the child's body, let alone the myriad of forces between atoms in the objects, but by doing so, we can easily solve some very complex problems with only minimal error due to our simplification.

Now, it also seems reasonable that acceleration should be inversely proportional to the mass of the system. In other words, the larger the mass (the inertia), the smaller the acceleration produced by a given force. And indeed, as illustrated in Figure 4.4, the same net external force applied to a car produces a much smaller acceleration than when applied to a basketball. The proportionality is written as

$$\mathbf{a} \propto (1/m) \text{ --- } \textbf{Equation 2}$$

where m is the mass of the system. Experiments have shown that acceleration is exactly inversely proportional to mass, just as it is exactly linearly proportional to the net external force.

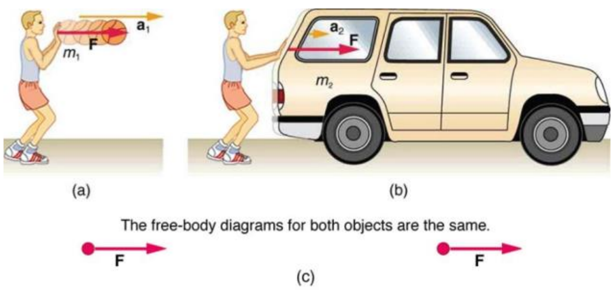


Figure **4.4:** The same force exerted on systems of different masses produces different accelerations. (a) A basketball player pushes on a basketball to make a pass. (The effect of gravity on the ball is ignored.) (b) The same player exerts an identical force on a stalled SUV and produces a far smaller acceleration (even if friction is negligible). (c) The free-body diagrams are identical, permitting direct comparison of the two situations. A series of patterns for the free-body diagram will emerge as you do more problems.

It has been found that the acceleration of an object depends *only* on the net external force and the mass of the object. Combining the two proportionalities just given yields Newton's second law of motion.

4.3 Newton's Second Law of Motion

The acceleration of a system is directly proportional to and in the same direction as the net external force acting on the system, and inversely proportional to its mass.

In equation form, Newton's second law of motion is

$$\mathbf{a} = \mathbf{F}_{net} / m \text{ --- } \textbf{Equation 3}$$

This is often written in the more familiar form, when only the magnitude of force and acceleration are considered,

$$\mathbf{F}_{net} = m\mathbf{a} \text{ --- } \textbf{Equation 4}$$

Although these last two equations are really the same, the first gives more insight into what Newton's second law means. The law is a *cause and effect relationship* among three quantities that is not simply based on their definitions. The validity of the second law is completely based on experimental verification.

Units of Force

$\mathbf{F}_{net} = m\mathbf{a}$ is used to define the units of force in terms of the three basic units for mass, length, and time. The SI unit of force is called the newton (abbreviated N) and is the force needed to accelerate a 1-kg system at the rate of 1m/s2. That is, since $\mathbf{F}_{net}=m\mathbf{a}$,

$$1\ \text{N}=1\ \text{kg}\cdot\text{m/s}^2$$

While almost the entire world uses the newton for the unit of force, in the United States the most familiar unit of force is the pound (lb), where 1 N = 0.225 lb.

4.4 Weight and the Gravitational Force

When an object is dropped, it accelerates toward the center of Earth. Newton's second law states that a net force on an object is responsible for its acceleration. If air resistance is negligible, the net force on a falling object is the gravitational force, commonly called its weight **w**. Weight can be denoted as a vector **w** because it has a direction; *down* is, by definition, the direction of gravity, and hence weight is a downward force. The magnitude of weight is denoted as w. Galileo was instrumental in showing that, in the absence of air resistance, all objects fall with the same acceleration g. Using Galileo's result and Newton's second law, we can derive an equation for weight.

Consider an object with mass m falling downward toward Earth. It experiences only the downward force of gravity, which has magnitude w. Newton's second law states that the magnitude of the net external force on an object is $F_{net}=ma$.

Since the object experiences only the downward force of gravity, $F_{net}=w$. We know that the acceleration of an object due to gravity is g, or $a=g$. Substituting these into Newton's second law gives

Weight

This is the equation for *weight*—the gravitational force on a mass m:

$$w = mg \quad \text{--- } \textbf{Equation 5}$$

Since g=9.80m/s2 on Earth, the weight of a 1.0 kg object on Earth is 9.8 N, as we see:

$$w = mg = (1.0\ \text{kg})(9.80\text{m/s}^2) = 9.8\text{N}.$$

Recall that g can take a positive or negative value, depending on the positive direction in the coordinate system. Be sure to take this into consideration when solving problems with weight.

When the net external force on an object is its weight, we say that it is in free-fall. That is, the only force acting on the object is the force of gravity. In the real world, when objects fall downward toward Earth, they are never truly in free-fall because there is always some upward force from the air acting on the object.

The acceleration due to gravity g varies slightly over the surface of Earth, so that the weight of an object depends on location and is not an intrinsic property of the object. Weight varies dramatically if one leaves Earth's surface. On the Moon, for example, the acceleration due to gravity is only 1.625m/s2. A 1.0-kg mass thus has a weight of 9.8 N on Earth and only about 1.7 N on the Moon.

The broadest definition of weight in this sense is that *the weight of an object is the gravitational force on it from the nearest large body*, such as Earth, the Moon, the Sun, and so on. This is the most common and useful definition of weight in physics. It differs dramatically, however, from the definition of weight used by NASA and the popular media in relation to space travel and exploration. When they speak of “weightlessness” and “microgravity,” they are really referring to the phenomenon we call “free-fall” in physics. We shall use the above definition of weight, and we will make careful distinctions between free-fall and actual weightlessness.

It is important to be aware that weight and mass are very different physical quantities, although they are closely related. Mass is the quantity of matter (how much “stuff”) and does not vary in classical physics, whereas weight is the gravitational force and does vary depending on gravity. It is tempting to equate the two, since most of our examples take place on Earth, where the weight of an object only varies a little with the location of the object. Furthermore, the terms *mass* and *weight* are used interchangeably in everyday language; for example, our medical records often show our “weight” in kilograms, but never in the correct units of newtons.

4.5 Common Misconceptions: Mass vs. Weight

Mass and weight are often used interchangeably in everyday language. However, in science, these terms are distinctly different from one another. Mass is a measure of how much matter is in an object. The typical measure of mass is the kilogram (or the “slug” in English units). Weight, on the other hand, is a measure of the force of gravity acting on an object. Weight is equal to the mass of an object (m) multiplied by the acceleration due to gravity (g). Like any other force, weight is measured in terms of newtons (or pounds in English units).

Assuming the mass of an object is kept intact, it will remain the same, regardless of its location. However, because weight depends on the acceleration due to gravity, the weight of an object *can change* when the object enters into a region with stronger or weaker gravity. For example, the acceleration due to gravity on the Moon is $1.625 m/s^2$ (which is much less than the acceleration due to gravity on Earth, $9.80 m/s^2$). If you measured your weight on Earth and then measured your weight on the Moon, you would find that you “weigh” much less, even though you do not look any skinnier. This is because the force of gravity is weaker on the Moon. In fact, when people say that they are “losing weight,” they really mean that they are losing “mass” (which in turn causes them to weigh less).

4.6 Newton’s Third Law of Motion

Whenever one body exerts a force on a second body, the first body experiences a force that is equal in magnitude and opposite in direction to the force that it exerts.

This law represents a certain *symmetry in nature*: Forces always occur in pairs, and one body cannot exert a force on another without experiencing a force itself. We sometimes refer to this law loosely as “action-reaction,” where the force exerted is the action and the force experienced as a consequence is the reaction. Newton's third law has practical uses in analyzing the origin of forces and understanding which forces are external to a system.

We can readily see Newton's third law at work by taking a look at how people move about. Consider a swimmer pushing off from the side of a pool, as illustrated in Figure 4.5. She pushes against the pool wall with her feet and accelerates in the direction *opposite* to that of her push. The wall has exerted an equal and opposite force back on the swimmer. You might think that two equal and opposite forces would cancel, but they do not *because they act on different systems*. In this case, there are two systems that we could investigate: the swimmer or the wall. If we select the swimmer to be the system of interest, as in the figure, then $\mathbf{F}_{\text{wall}}$ on feet is an external force on this system and affects its motion. The swimmer moves in the direction of $\mathbf{F}_{\text{wall}}$ on feet. In contrast, the force $\mathbf{F}_{\text{feet}}$ on wall acts on the wall and not on our system of interest. Thus $\mathbf{F}_{\text{feet}}$ on wall does not directly affect the motion of the system and does not cancel $\mathbf{F}_{\text{wall}}$ on feet. Note that the swimmer pushes in the direction opposite to that in which she wishes to move. The reaction to her push is thus in the desired direction.

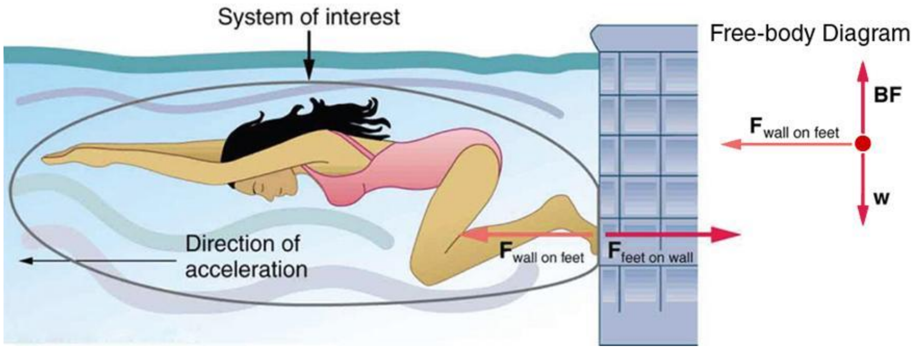


Figure **4.5:** When the swimmer exerts a force $\mathbf{F}_{\text{feet}}$ on wall on the wall, she accelerates in the direction opposite to that of her push. This means the net external force on her is in the direction opposite to $\mathbf{F}_{\text{feet}}$ on wall. This opposition occurs because, in accordance with Newton's third law of motion, the wall exerts a force $\mathbf{F}_{\text{wall}}$ on feet on her, equal in magnitude but in the direction opposite to the one she exerts on it. The line around the swimmer indicates the system of interest. Note that $\mathbf{F}_{\text{feet}}$ on wall does not act on this system (the swimmer) and, thus, does not cancel $\mathbf{F}_{\text{wall}}$ on feet. Thus the free-body diagram shows only $\mathbf{F}_{\text{wall}}$ on feet, **w**, the gravitational force, and **BF**, the buoyant force of the water supporting the swimmer's weight. The vertical forces **w** and **BF** cancel since there is no vertical motion.

Other examples of Newton's third law are easy to find. As a professor walks in front of a whiteboard, she exerts a force backward on the floor. The floor exerts a reaction force forward on the professor that causes her to accelerate forward. Similarly, a car accelerates because the ground pushes forward on the drive wheels in reaction to the drive wheels pushing backward on the ground. You can see evidence of the wheels pushing backward when tires spin on a gravel road and throw rocks backward. In another example, rockets move forward by expelling gas backward at high velocity. This means the rocket exerts a large backward force on the gas in the rocket combustion chamber, and the gas therefore exerts a large reaction force forward on the rocket. This reaction force is called thrust. It is a common misconception that rockets propel themselves by pushing on the ground or on the air behind them. They actually work better in a vacuum, where they can more readily expel the exhaust gases. Helicopters similarly create lift by pushing air down, thereby experiencing an upward reaction force. Birds and airplanes also fly by exerting force on air in a direction opposite to that of whatever force they need. For example, the wings of a bird force air downward and backward in order to get lift and move forward. An octopus propels itself in the water by ejecting water through a funnel from its body, similar to a jet ski. Boxers and other martial arts fighters experience reaction forces when they punch, sometimes breaking their hand by hitting an opponent's body.

Example:

A physics professor pushes a cart of demonstration equipment to a lecture hall, as seen in Figure 4.6. Her mass is 65.0 kg, the cart's is 12.0 kg, and the equipment's is 7.0 kg. Calculate the acceleration produced when the professor exerts a backward force of 150 N on the floor. All forces opposing the motion, such as friction on the cart's wheels and air resistance, total 24.0 N.

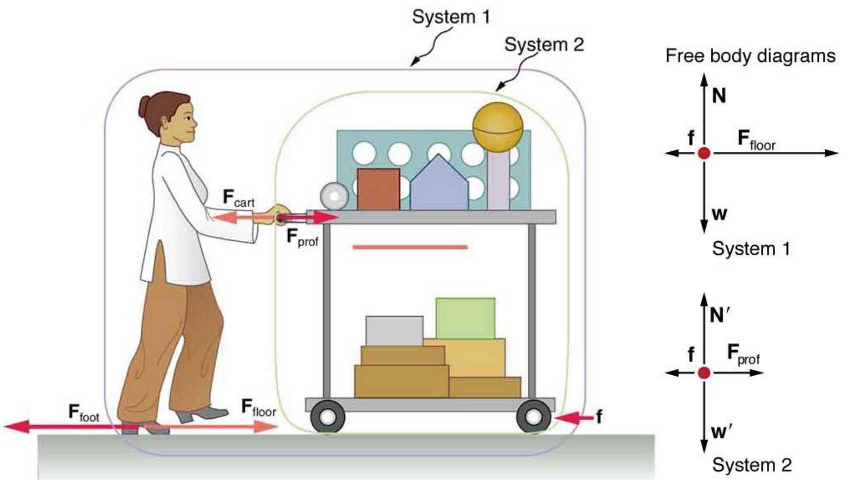


Figure **4.6:** A professor pushes a cart of demonstration equipment. The lengths of the arrows are proportional to the magnitudes of the forces (except for **f**, since it is too small to draw to scale). Different questions are asked in each example; thus, the system of interest must be defined differently for each. System 1 is appropriate for this example, since it asks for the acceleration of the entire group of objects. Only $\mathbf{F}_{floor}$ and **f** are external forces acting on System 1 along the line of motion. All other forces either cancel or act on the outside world. System 2 is chosen, so that **F**prof will be an external force and enter into Newton's second law. Note that the free-body diagrams, which allow us to apply Newton's second law, vary with the system chosen.

Strategy

Since they accelerate as a unit, we define the system to be the professor, cart, and equipment. This is System 1 in Figure 4.6. The professor pushes backward with a force $\mathbf{F}_{\text{foot}}$ of 150 N. According to Newton's third law, the floor exerts a forward reaction force $\mathbf{F}_{\text{floor}}$ of 150 N on System 1. Because all motion is horizontal, we can assume there is no net force in the vertical direction. The problem is therefore one-dimensional along the horizontal direction. As noted, **f** opposes the motion and is thus in the opposite direction of $\mathbf{F}_{\text{floor}}$. Note that we do not include the forces $\mathbf{F}_{\text{prof}}$ or $\mathbf{F}_{\text{cart}}$ because these are internal forces, and we do not include **F**foot because it acts on the floor, not on the system. There are no other significant forces acting on System 1. If the net external force can be found from all this information, we can use Newton's second law to find the acceleration as requested. See the free-body diagram in the figure.

Solution

Newton's second law is given by,

$$a = F_{\text{net}}\, m$$

The net external force on System 1 is deduced from Figure 4.6 and the discussion above to be

$$F\text{net} = F\text{floor} - f = 150\text{ N} - 24.0\text{ N} = 126\text{ N}$$

The mass of System 1 is,

$$m = (65.0+12.0+7.0)\text{kg} = 84\text{ kg}$$

These values of Fnet and m produce an acceleration of

$$a = F\text{net}\ m$$

$$a = 126\text{ N X } 84\text{ kg} = 1.5\text{ m/s}^2$$

Discussion

None of the forces between components of System 1, such as between the professor's hands and the cart, contribute to the net external force because they are internal to System 1. Another way to look at this is to note that forces between components of a system cancel because they are equal in magnitude and opposite in direction. For example, the force exerted by the professor on the cart results in an equal and opposite force back on her. In this case both forces act on the same system and, therefore, cancel. Thus internal forces (between components of a system) cancel. Choosing System 1 was crucial to solving this problem.

Example: Force on the Cart—Choosing a New System

Calculate the force the professor exerts on the cart in Figure 4.6 using data from the previous example if needed.

Strategy

If we now define the system of interest to be the cart plus equipment (System 2 in Figure 4.6), then the net external force on System 2 is the force the professor exerts on the cart minus friction. The force she exerts on the cart, $\mathbf{F}_{prof}$, is an external force acting on System 2. $\mathbf{F}_{prof}$ was internal to System 1, but it is external to System 2 and will enter Newton's second law for System 2.

Solution

Newton's second law can be used to find $\mathbf{F}_{prof}$. Starting with

$$a = F_{net}\ m$$

and noting that the magnitude of the net external force on System 2 is

$$F_{net} = F_{prof} - f$$

we solve for F_{prof}, the desired quantity:

$$F_{prof} = F_{net} + f$$

The value of f is given, so we must calculate net Fnet. That can be done since both the acceleration and mass of System 2 are known. Using Newton's second law we see that ,

$$F_{net} = ma$$

where the mass of System 2 is 19.0 kg (m = 12.0 kg + 7.0 kg) and its acceleration was found to be $a = 1.5\text{m/s}^2$ in the previous example. Thus,

$$F_{net} = ma$$

$$F_{net} = (19.0\text{ kg})(1.5\text{m/s}^2) = 29\text{ N}$$

Now we can find the desired force:

$$F_{prof} = F_{net} + f$$

$$F_{prof} = 29\text{ N}+24.0\text{ N} = 53\text{ N}$$

Discussion

It is interesting that this force is significantly less than the 150N force the professor exerted backward on the floor. Not all of that 150N force is transmitted to the cart; some of it accelerates the professor.

The choice of a system is an important analytical step both in solving problems and in thoroughly understanding the physics of the situation (which is not necessarily the same thing).

4.7 Normal Force

Weight (also called force of gravity) is a pervasive force that acts at all times and must be counteracted to keep an object from falling. You definitely notice that you must support the weight of a heavy object by pushing up on it when you hold it stationary, as illustrated in Figure 4.7 (a). But how do inanimate objects like a table support the weight of a mass placed on them, such as shown in Figure 4.7 (b)? When the bag of dog food is placed on the table, the table actually sags slightly under the load. This would be noticeable if the load were placed on a card table, but even rigid objects deform when a force is applied to them. Unless the object is deformed beyond its limit, it will exert a restoring force much like a deformed spring (or trampoline or diving board). The greater the deformation, the greater the restoring force. So, when the load is placed on the table, the table sags until the restoring force becomes as large as the weight of the load. At this point the net external force on the load is zero. That is the situation when the load is stationary on the table. The table sags quickly, and the sag is slight so we do not notice it. But it is similar to the sagging of a trampoline when you climb onto it.

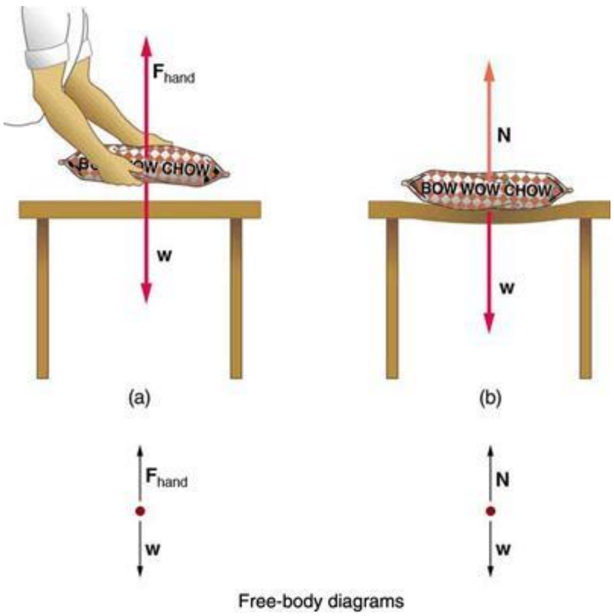


Figure **4.7:** (a) The person holding the bag of dog food must supply an upward force **F**hand equal in magnitude and opposite in direction to the weight of the food **w**. (b) The card table sags when the dog food is placed on it, much like a stiff trampoline. Elastic restoring forces in the table grow as it sags until they supply a force **N** equal in magnitude and opposite in direction to the weight of the load.

We must conclude that whatever supports a load, be it animate or not, must supply an upward force equal to the weight of the load, as we assumed in a few of the previous examples. If the force supporting a load is perpendicular to the surface of contact between the load and its support, this force is defined to be a normal force and here is given the symbol **N**. (This is not the unit for force N.) The word *normal* means perpendicular to a surface. The normal force can be less than the object’s weight if the object is on an incline, as you will see in the next example.

Common Misconception: Normal Force (N) vs. Newton (N)

In this section we have introduced the quantity normal force, which is represented by the variable **N**. This should not be confused with the symbol for the newton, which is also represented by the letter N. These symbols are particularly important to distinguish because the units of a normal force (**N**) happen to be newtons (N). For example, the normal force **N** that the floor exerts on a chair might be **N**=100 N. One important difference is that normal force is a vector, while the newton is simply a unit. Be careful not to confuse these letters in your calculations! You will encounter more similarities among variables and units as you proceed in physics. Another example of this is the quantity work (*W*) and the unit watts (W).

Example - Weight on an Incline, a Two-Dimensional Problem

Consider the skier on a slope shown in Figure 4.8. Her mass including equipment is 60.0 kg. (a) What is her acceleration if friction is negligible? (b) What is her acceleration if friction is known to be 45.0 N?

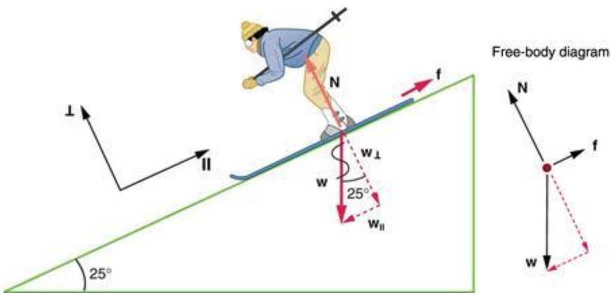


Figure **4.8:** Since motion and friction are parallel to the slope, it is most convenient to project all forces onto a coordinate system where one axis is parallel to the slope and the other is perpendicular (axes shown to left of skier). **N** is perpendicular to the slope and **f** is parallel to the slope, but **w** has components along both axes, namely **w**⊥ and **w**∥. **N** is equal in magnitude to **w**⊥, so that there is no motion perpendicular to the slope, but *f* is less than *w*∥, so that there is a downslope acceleration (along the parallel axis).

Strategy

This is a two-dimensional problem, since the forces on the skier (the system of interest) are not parallel. The approach we have used in two-dimensional kinematics also works very well here. Choose a convenient coordinate system and project the vectors onto its axes, creating *two* connected *one*-dimensional problems to solve. The most convenient coordinate system for motion on an incline is one that has one coordinate parallel to the slope and one perpendicular to the slope. (Remember that motions along mutually perpendicular axes are independent.) We use the symbols ⊥ and ∥ to represent perpendicular and parallel, respectively. This choice of axes simplifies this type of problem, because there is no motion perpendicular to the slope and because friction is always parallel to the surface between two objects. The only external forces acting on the system are the skier's weight, friction, and the support of the slope, respectively labeled **w**, **f**, and **N** in Figure 4.8. **N** is always perpendicular to the slope, and **f** is parallel to it. But **w** is not in the direction of either axis, and so the first step we take is to project it into components along the chosen axes, defining $w_{\parallel}$ to be the component of weight parallel to the slope and $w_{\perp}$ the component of weight perpendicular to the slope. Once this is done, we can consider the two separate problems of forces parallel to the slope and forces perpendicular to the slope.

Solution

The magnitude of the component of the weight parallel to the slope is,

$$w_{\parallel} = w\sin(25^{\circ}) = mg\sin(25^{\circ}),$$

and the magnitude of the component of the weight perpendicular to the slope is,

$$w_{\perp} = w\cos(25^{\circ}) = mg\cos(25^{\circ}).$$

(a) Neglecting friction. Since the acceleration is parallel to the slope, we need only consider forces parallel to the slope. (Forces perpendicular to the slope add to zero, since there is no acceleration in that direction.) The forces parallel to the slope are the amount of the skier's weight parallel to the slope $w_{\parallel}$ and friction f. Using Newton's second law, with subscripts to denote quantities parallel to the slope,

$$a_{\parallel} = F_{net\parallel}\ m$$

where $F_{net\parallel}=w_{\parallel}=mg\sin(25^{\circ})$, assuming no friction for this part, so that

$$a_{\parallel} = F_{net\parallel}\ m = mg\sin(25^{\circ})m = g\sin(25^{\circ})$$

$$(9.80 \text{m/s}^2)(0.4226) = 4.14 \text{m/s}^2$$

This is the acceleration.

(b) Including friction. We now have a given value for friction, and we know its direction is parallel to the slope and it opposes motion between surfaces in contact. So the net external force is now

$$F_{net\parallel} = w_{\parallel} - f$$

and substituting this into Newton's second law, $a_{\parallel}=F_{net\parallel}\ m$, gives

$$a_{\parallel} = F_{net||}\ m = w_{\parallel} - fm = mg\sin(25^{\circ}) - fm.$$

We substitute known values to obtain

$$a_{\parallel} = (60.0 \text{ kg})(9.80 \text{ m/s}^2)(0.4226) - 45.0 \text{ N } 60.0 \text{ kg},$$

which yields,

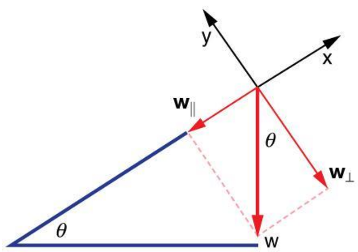
$$a_{\parallel} = 3.39 \text{ m/s}^2$$

This is the acceleration parallel to the incline when there is 45.0 N of opposing friction.

Discussion

Since friction always opposes motion between surfaces, the acceleration is smaller when there is friction than when there is none. In fact, it is a general result that if friction on an incline is negligible, then the acceleration down the incline is $a = g\sin\theta$, *regardless of mass*. This is related to the previously discussed fact that all objects fall with the same acceleration in the absence of air resistance. Similarly, all objects, regardless of mass, slide down a frictionless incline with the same acceleration (if the angle is the same).

Resolving Weight into Components



$$w_{\parallel} = w\sin(\theta) = mg\sin(\theta)$$
$$w_{\perp} = w\cos(\theta) = mg\cos(\theta)$$

Figure **4.9:** An object rests on an incline that makes an angle θ with the horizontal.

When an object rests on an incline that makes an angle θ with the horizontal, the force of gravity acting on the object is divided into two components: a force acting perpendicular to the plane, **w⊥,** and a force acting parallel to the plane, **w‖**. The perpendicular force of weight, **w⊥,** is typically equal in magnitude and opposite in direction to the normal force, **N**. The force acting parallel to the plane, **w‖**, causes the object to accelerate down the incline. The force of friction, **f**, opposes the motion of the object, so it acts upward along the plane.

It is important to be careful when resolving the weight of the object into components. If the angle of the incline is at an angle ϑ to the horizontal, then the magnitudes of the weight components are,

$$w_{\parallel} = w\sin(\theta) = mg\sin(\theta)$$

and

$$w_{\perp} = w\cos(\theta) = mg\cos(\theta)$$

Instead of memorizing these equations, it is helpful to be able to determine them from reason. To do this, draw the right triangle formed by the three weight vectors. Notice that the angle ϑ of the incline is the same as the angle formed between **w** and **w⊥**. Knowing this property, you can use trigonometry to determine the magnitude of the weight components:

$$\cos(\theta) = w_{\perp} / w$$

$$w_{\perp} = w\cos(\theta) = mg\cos(\theta)$$

$$\sin(\theta) = w_{\parallel} / w$$

$$w_{\parallel} = w\sin(\theta) = mg\sin(\theta)$$

4.8 Tension

A tension is a force along the length of a medium, especially a force carried by a flexible medium, such as a rope or cable. The word "tension" comes from a Latin word meaning "to stretch." Not coincidentally, the flexible cords that carry muscle forces to other parts of the body are called *tendons*. Any flexible connector, such as a string, rope, chain, wire, or cable, can exert pulls only parallel to its length; thus, a force carried by a flexible connector is a tension with direction parallel to the connector. It is important to understand that tension is a pull in a connector. In contrast, consider the phrase: "You can't push a rope." The tension force pulls outward along the two ends of a rope.

Consider a person holding a mass on a rope as shown in Figure 4.10.

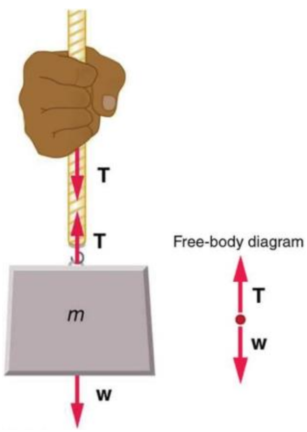


Figure **4.10:** When a perfectly flexible connector (one requiring no force to bend it) such as this rope transmits a force **T**, that force must be parallel to the length of the rope, as shown. The pull such a flexible connector exerts is a tension. Note that the rope pulls with equal force but in opposite directions on the hand and the supported mass (neglecting the weight of the rope). This is an example of Newton's third law. The rope is the medium that carries the equal and opposite forces between the two objects. The tension anywhere in the rope between the hand and the mass is equal. Once you have determined the tension in one location, you have determined the tension at all locations along the rope.

Tension in the rope must equal the weight of the supported mass, as we can prove using Newton's second law. If the 5.00-kg mass in the figure is stationary, then its acceleration is zero, and thus $\mathbf{F}_{\text{net}}=0$. The only external forces acting on the mass are its weight $\mathbf{w}$ and the tension $\mathbf{T}$ supplied by the rope. Thus,

$$F_{\text{net}} = T - w = 0$$

where T and w are the magnitudes of the tension and weight and their signs indicate direction, with up being positive here. Thus, just as you would expect, the tension equals the weight of the supported mass:

$$T = w = mg$$

For a 5.00-kg mass, then (neglecting the mass of the rope) we see that,

$$T = mg = (5.00\text{ kg})(9.80\text{ m/s}^2) = 49.0\text{ N}$$

If we cut the rope and insert a spring, the spring would extend a length corresponding to a force of 49.0 N, providing a direct observation and measure of the tension force in the rope.

Flexible connectors are often used to transmit forces around corners, such as in a hospital traction system, a finger joint, or a bicycle brake cable. If there is no friction, the tension is transmitted undiminished. Only its direction changes, and it is always parallel to the flexible connector. This is illustrated in Figure 4.11 (a) and (b).

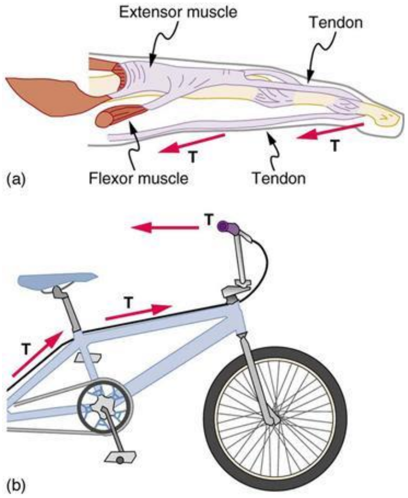


Figure **4.11:** (a) Tendons in the finger carry force **T** from the muscles to other parts of the finger, usually changing the force's direction, but not its magnitude (the tendons are relatively friction free). (b) The brake cable on a bicycle carries the tension **T** from the handlebars to the brake mechanism. Again, the direction but not the magnitude of **T** is changed.

Example - the Tension in a Tightrope?

Calculate the tension in the wire supporting the 70.0-kg tightrope walker shown in Figure 4.12.

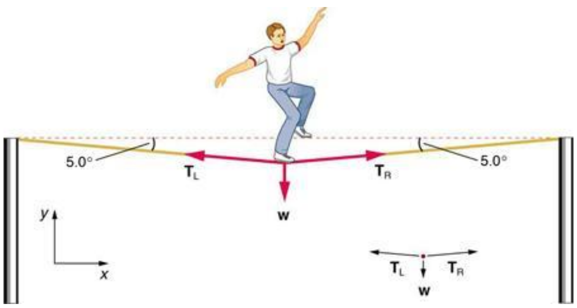


Figure **4.12:** The weight of a tightrope walker causes a wire to sag by 5.0 degrees. The system of interest here is the point in the wire at which the tightrope walker is standing.

Strategy

As you can see in the figure, the wire is not perfectly horizontal (it cannot be!), but is bent under the person's weight. Thus, the tension on either side of the person has an upward component that can support his weight. As usual, forces are vectors represented pictorially by arrows having the same directions as the forces and lengths proportional to their magnitudes. The system is the tightrope walker, and the only external forces acting on him are his weight $\mathbf{w}$ and the two tensions $\mathbf{T}_L$ (left tension) and $\mathbf{T}_R$ (right tension), as illustrated. It is reasonable to neglect the weight of the wire itself. The net external force is zero since the system is stationary. A little trigonometry can now be used to find the tensions. One conclusion is possible at the outset—we can see from part (b) of the figure that the magnitudes of the tensions T_L and T_R must be equal. This is because there is no horizontal acceleration in the rope, and the only forces acting to the left and right are T_L and T_R. Thus, the magnitude of those forces must be equal so that they cancel each other out.

Whenever we have two-dimensional vector problems in which no two vectors are parallel, the easiest method of solution is to pick a convenient coordinate system and project the vectors onto its axes. In this case the best coordinate system has one axis horizontal and the other vertical. We call the horizontal the x-axis and the vertical the y-axis.

Solution

First, we need to resolve the tension vectors into their horizontal and vertical components. It helps to draw a new free-body diagram showing all of the horizontal and vertical components of each force acting on the system.

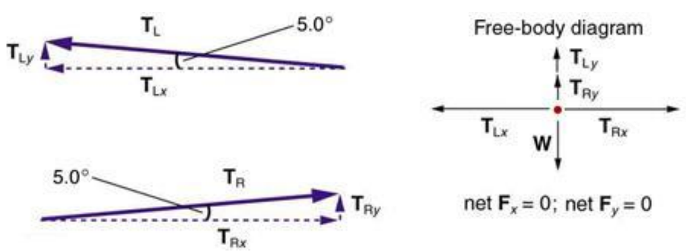


Figure **4.13**: When the vectors are projected onto vertical and horizontal axes, their components along those axes must add to zero, since the tightrope walker is stationary. The small angle results in *T* being much greater than *w*.

Consider the horizontal components of the forces (denoted with a subscript x):

$$F_{netx} = T_{Lx} - T_{Rx}$$

The net external horizontal force $F_{netx} = 0$, since the person is stationary. Thus,

$$F_{netx} = 0 = T_{Lx} - T_{Rx}$$

$$T_{Lx} = T_{Rx}$$

Now, observe Figure 4.13. You can use trigonometry to determine the magnitude of T_L and T_R. Notice that:

$$\cos(5.0^o) = T_{Lx} / T_L$$

$$T_{Lx} = T_L \cos(5.0^o)$$

$$\cos(5.0^o) = T_{Rx} / T_R$$

$$T_{Rx} = T_R \cos(5.0^0)$$

Equating T_{Lx} and T_{Rx}:

$$T_L \cos(5.0^o) = T_R \cos(5.0^o)$$

Thus,

$$T_L = T_R = T$$

as predicted. Now, considering the vertical components (denoted by a subscript *y*), we can solve for *T*. Again, since the person is stationary, Newton's second law implies that net Fy=0. Thus, as illustrated in the free-body diagram in Figure 4.13,

$$F_{net\ y} = T_{Ly} + T_{ry} - w = 0$$

Observing Figure 4.13, we can use trigonometry to determine the relationship between T_{Ly}, T_{Ry}, and T. As we determined from the analysis in the horizontal direction, $T_L = T_R = T$:

$$\sin(5.0^o) = T_{Ly} / T_L$$

$$T_{Ly} = T_L\sin(5.0^o) = T\sin(5.0^0)$$

$$\sin(5.0^o) = T_{Ry} / T_R$$

$$T_{Ry} = T_R\sin(5.0^o) = T\sin(5.0^o).$$

Now, we can substitute the values for T_{Ly} and T_{Ry}, into the net force equation in the vertical direction:

$$F_{nety} = T_{Ly} + T_{Ry} – w = 0$$

$$F_{nety} = T\sin(5.0^o) + T\sin(5.0^o) – w = 0$$

$$2T\sin(5.0^o) – w = 0$$

$$2T\sin(5.0^o) = w$$

and

$$T = w / (2\sin(5.0^o)) = mg / (2\sin(5.0^o))$$

so that

$$T = (70.0\ \text{kg})(9.80\ \text{m/s}^2) / (2(0.0872))$$

and the tension is

$$T = 3900\ \text{N}$$

Discussion

Note that the vertical tension in the wire acts as a normal force that supports the weight of the tightrope walker. The tension is almost six times the 686 N weight of the tightrope walker. Since the wire is nearly horizontal, the vertical component of its tension is only a small fraction of the tension in the wire. The large horizontal components are in opposite directions and cancel, and so most of the tension in the wire is not used to support the weight of the tightrope walker.

If we wish to *create* a very large tension, all we have to do is exert a force perpendicular to a flexible connector, as illustrated in Figure 4.14. As we saw in the last example, the weight of the tightrope walker acted as a force perpendicular to the rope. We saw that the tension in the roped related to the weight of the tightrope walker in the following way:

$$T = w / (2\sin(\theta))$$

We can extend this expression to describe the tension T created when a perpendicular force ($\mathbf{F}\perp$) is exerted at the middle of a flexible connector:

$$T = F\perp / (2\sin(\theta))$$

Note that θ is the angle between the horizontal and the bent connector. In this case, T becomes very large as θ approaches zero. Even the relatively small weight of any flexible connector will cause it to sag, since an infinite tension would result if it were horizontal (i.e., $\theta = 0$ and $\sin\theta = 0$). (See Figure 4.14)

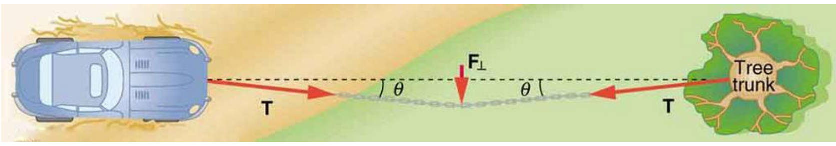


Figure **4.18:** We can create a very large tension in the chain by pushing on it perpendicular to its length, as shown. Suppose we wish to pull a car out of the mud when no tow truck is available. Each time the car moves forward, the chain is tightened to keep it as nearly straight as possible. The tension in the chain is given by $T = F\perp \,/\, (2\sin(\theta))$; since θ is small, T is very large. This situation is analogous to the tightrope walker shown in Figure 4.12, except that the tensions shown here are those transmitted to the car and the tree rather than those acting at the point where $\mathbf{F}\perp$ is applied.

4.9 Real Forces and Inertial Frames

There is another distinction among forces in addition to the types already mentioned. Some forces are real, whereas others are not. *Real forces* are those that have some physical origin, such as the gravitational pull. Contrastingly, *fictitious forces* are those that arise simply because an observer is in an accelerating frame of reference, such as one that rotates (like a merry-go-round) or undergoes linear acceleration (like a car slowing down). For example, if a satellite is heading due north above Earth's northern hemisphere, then to an observer on Earth it will appear to experience a force to the west that has no physical origin. Of course, what is happening here is that Earth is rotating toward the east and moves east under the satellite. In Earth's frame this looks like a westward force on the satellite, or it can be interpreted as a violation of Newton's first law (the law of inertia). An inertial frame of reference is one in which all forces are real and, equivalently, one in which Newton's laws have the simple forms given in this chapter.

Earth's rotation is slow enough that Earth is nearly an inertial frame. You ordinarily must perform precise experiments to observe fictitious forces and the slight departures from Newton's laws, such as the effect just described. On the large scale, such as for the rotation of weather systems and ocean currents, the effects can be easily observed.

The crucial factor in determining whether a frame of reference is inertial is whether it accelerates or rotates relative to a known inertial frame. Unless stated otherwise, all phenomena discussed in this text are considered in inertial frames.

All the forces discussed in this section are real forces, but there are a number of other real forces, such as lift and thrust, that are not discussed in this section. They are more specialized, and it is not necessary to discuss every type of force.

The Four Basic Forces

The four basic forces will be encountered in more detail as you progress in this book. The gravitational force is defined in Uniform Circular Motion and Gravitation, electric force in Electric Charge and Electric Field, magnetic force in Magnetism, and nuclear forces in Radioactivity and Nuclear Physics. On a macroscopic scale, electromagnetism and gravity are the basis for all forces. The nuclear forces are vital to the substructure of matter, but they are not directly experienced on the macroscopic scale.

Force	Approximate Relative Strengths	Range	Attraction/Repulsion	Carrier Particle
Gravitational	10−38	∞	attractive only	Graviton
Electromagnetic	10–2	∞	attractive and repulsive	Photon
Weak nuclear	10–6	< 10–18m	attractive and repulsive	W+, W–, Z0
Strong nuclear	1	∞	attractive and repulsive	gluons

Table 4.1: Properties of the Four Basic Forces

The gravitational force is surprisingly weak—it is only because gravity is always attractive that we notice it at all. Our weight is the gravitational force due to the *entire* Earth acting on us. On the very large scale, as in astronomical systems, the gravitational force is the dominant force determining the motions of moons, planets, stars, and galaxies. The gravitational force also affects the nature of space and time. In the study of general relativity, space is considered curved in the vicinity of very massive bodies, such as the Sun, and time actually slows down near massive bodies.

Electromagnetic forces can be either attractive or repulsive. They are long-range forces, which act over extremely large distances, and they nearly cancel for macroscopic objects. (Remember that it is the *net* external force that is important.) If they did not cancel, electromagnetic forces would completely overwhelm the gravitational force. The electromagnetic force is a combination of electrical forces (such as those that cause static electricity) and magnetic forces (such as those that affect a compass needle). These two forces were thought to be quite distinct until early in the 19th century, when scientists began to discover that they are different manifestations of the same force. This discovery is a classical case of the *unification of forces*. Similarly, friction, tension, and all of the other classes of forces we experience directly (except gravity, of course) are due to electromagnetic interactions of atoms and molecules. It is still convenient to consider these forces separately in specific applications, however, because of the ways they manifest themselves.

For examples and answers, please refer to OpenStax.com questions and answers given on their website or to the College Physics 2e - https://openstax.org/details/books/college-physics-2e

Chapter 5

5.0 Objectives

At the end of this lesson, students should be able to,

- Identify friction.
- Calculate friction.
- Calculate drag force.
- Identify stress and strain.
- Calculate deformation.

5.1 Introduction

This chapter covers friction, drag force, stress, strain, deformation and relevant calculations.

It is difficult to categorize forces into various types (aside from the four basic forces discussed in a previous chapter). We know that a net force affects the motion, position, and shape of an object. It is useful at this point to look at some particularly interesting and common forces that will provide further applications of Newton's laws of motion. We have in mind the forces of friction, air or liquid drag, and deformation.

Friction is a force that is around us all the time that opposes relative motion between surfaces in contact but also allows us to move (which you have discovered if you have ever tried to walk on ice). While a common force, the behavior of friction is actually very complicated and is still not completely understood. We have to rely heavily on observations for whatever understandings we can gain. However, we can still deal with its more elementary general characteristics and understand the circumstances in which it behaves.

5.2 Friction

Friction is a force that opposes relative motion between surfaces in contact.

One of the simpler characteristics of friction is that it is parallel to the contact surface between surfaces and always in a direction that opposes motion or attempted motion of the systems relative to each other. If two surfaces are in contact and moving relative to one another, then the friction between them is called kinetic friction. For example, friction slows a hockey puck sliding on ice. But when objects are stationary, static friction can act between them; the static friction is usually greater than the kinetic friction between the surfaces.

Kinetic Friction

If two surfaces are in contact and moving relative to one another, then the friction between them is called kinetic friction.

Imagine, for example, trying to slide a heavy crate across a concrete floor—you may push harder and harder on the crate and not move it at all. This means that the static friction responds to what you do—it increases to be equal to and in the opposite direction of your push. But if you finally push hard enough, the crate seems to slip suddenly and starts to move. Once in motion it is easier to keep it in motion than it was to get it started, indicating that the kinetic friction force is less than the static friction force. If you add mass to the crate, say by placing a box on top of it, you need to push even harder to get it started and also to keep it moving. Furthermore, if you oiled the concrete you would find it to be easier to get the crate started and keep it going (as you might expect).

Figure 5.1 is a crude pictorial representation of how friction occurs at the interface between two objects. Close-up inspection of these surfaces shows them to be rough. So, when you push to get an object moving (in this case, a crate), you must raise the object until it can skip along with just the tips of the surface hitting, break off the points, or do both. A considerable force can be resisted by friction with no apparent motion. The harder the surfaces are pushed together (such as if another box is placed on the crate), the more force is needed to move them. Part of the friction is due to adhesive forces between the surface molecules of the two objects, which explain the dependence of friction on the nature of the substances. Adhesion varies with substances in contact and is a complicated aspect of surface physics. Once an object is moving, there are fewer points of contact (fewer molecules adhering), so less force is required to keep the object moving. At small but nonzero speeds, friction is nearly independent of speed.

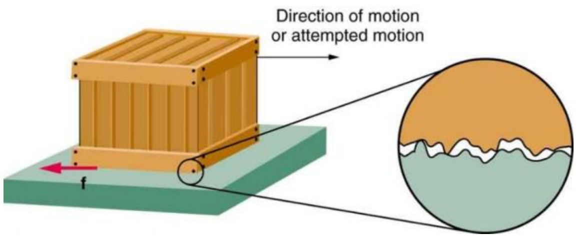


Figure 5.1: Frictional forces, such as *f*, always oppose motion or attempted motion between surfaces in contact. Friction arises in part because of the roughness of the surfaces in contact, as seen in the expanded view. In order for the object to move, it must rise to where the peaks can skip along the bottom surface. Thus, a force is required just to set the object in motion. Some of the peaks will be broken off, also requiring a force to maintain motion. Much of the friction is actually due to attractive forces between molecules making up the two objects, so that even perfectly smooth surfaces are not friction-free. Such adhesive forces also depend on the substances the surfaces are made of, explaining, for example, why rubber-soled shoes slip less than those with leather soles.

The magnitude of the frictional force has two forms: one for static situations (static friction), the other for when there is motion (kinetic friction).

When there is no motion between the objects, the magnitude of static friction $\mathbf{f}_s$ is

$$f_s \leq \mu_s N$$

where μ_s is the coefficient of static friction and N is the magnitude of the normal force (the force perpendicular to the surface).

Magnitude of Static Friction

Magnitude of static friction f_s is

$$f_s \leq \mu_s N,$$

where μ_s is the coefficient of static friction and N is the magnitude of the normal force.

The symbol ≤ means *less than or equal to*, implying that static friction can have a minimum and a maximum value of $\mu_s N$. Static friction is a responsive force that increases to be equal and opposite to whatever force is exerted, up to its maximum limit. Once the applied force exceeds $f_{s(max)}$, the object will move. Thus

$$f_{s(max)} = \mu_s N$$

Once an object is moving, the magnitude of kinetic friction $\mathbf{f}_k$ is given by

$$f_k = \mu_k N,$$

where μ_k is the coefficient of kinetic friction. A system in which $f_k = \mu_k N$ is described as a system in which *friction behaves simply*.

Magnitude of Kinetic Friction

The magnitude of kinetic friction f_k is given by

$$f_k = \mu_k N$$

where μ_k is the coefficient of kinetic friction.

As seen in Table 5.1, the coefficients of kinetic friction are less than their static counterparts. That values of μ in Table 5.1 are stated to only one or, at most, two digits is an indication of the approximate description of friction given by the above two equations.

System	**Static friction** μ_s	**Kinetic friction** μ_k
Rubber on dry concrete	1.0	0.7
Rubber on wet concrete	0.7	0.5
Wood on wood	0.5	0.3
Waxed wood on wet snow	0.14	0.1
Metal on wood	0.5	0.3

System	Static friction μ_s	Kinetic friction μ_k
Steel on steel (dry)	0.6	0.3
Steel on steel (oiled)	0.05	0.03
Teflon on steel	0.04	0.04
Bone lubricated by synovial fluid	0.016	0.015
Shoes on wood	0.9	0.7
Shoes on ice	0.1	0.05
Ice on ice	0.1	0.03
Steel on ice	0.04	0.02

Table **5.1:** Coefficients of Static and Kinetic Friction

The equations given earlier include the dependence of friction on materials and the normal force. The direction of friction is always opposite that of motion, parallel to the surface between objects, and perpendicular to the normal force. For example, if the crate you try to push (with a force parallel to the floor) has a mass of 100 kg, then the normal force would be equal to its weight,

$$W = mg = (100\text{ kg})(9.80\text{m/s}^2) = 980\text{ N},$$

perpendicular to the floor. If the coefficient of static friction is 0.45, you would have to exert a force parallel to the floor greater than,

$$f_s(\text{max}) = \mu_s N = (0.45)(980\text{N}) = 440\text{N}$$

to move the crate. Once there is motion, friction is less and the coefficient of kinetic friction might be 0.30, so that a force of only 290 N [$f_k = \mu_k N = (0.30)(980\text{N}) = 290\text{N}$]

would keep it moving at a constant speed. If the floor is lubricated, both coefficients are considerably less than they would be without lubrication. Coefficient of friction is a unit less quantity with a magnitude usually between 0 and 1.0. The coefficient of the friction depends on the two surfaces that are in contact.

Example - Skiing Exercise

A skier with a mass of 62 kg is sliding down a snowy slope. Find the coefficient of kinetic friction for the skier if friction is known to be 45.0 N.

Strategy

The magnitude of kinetic friction was given in to be 45.0 N. Kinetic friction is related to the normal force N as $f_k=\mu_k N$; thus, the coefficient of kinetic friction can be found if we can find the normal force of the skier on a slope. The normal force is always perpendicular to the surface, and since there is no motion perpendicular to the surface, the normal force should equal the component of the skier's weight perpendicular to the slope. (See the skier and free-body diagram in Figure 5.2.)

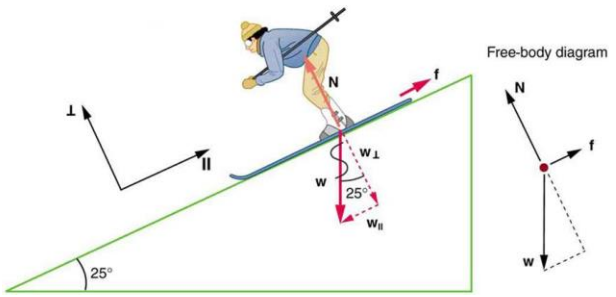


Figure **5.2:** The motion of the skier and friction are parallel to the slope and so it is most convenient to project all forces onto a coordinate system where one axis is parallel to the slope and the other is perpendicular (axes shown to left of skier). **N** (the normal force) is perpendicular to the slope, and **f** (the friction) is parallel to the slope, but **w** (the skier's weight) has components along both axes, namely **w**⊥ and **W**//. **N** is equal in magnitude to **w**⊥, so there is no motion perpendicular to the slope. However, **f** is less than **W**// in magnitude, so there is acceleration down the slope (along the *x*-axis).

That is,

$$N = w_{\perp} = w\cos 25^{\circ} = mg\cos 25^{\circ}$$

Substituting this into our expression for kinetic friction, we get

$$F_k = \mu_k mg\cos 25^{\circ}$$

which can now be solved for the coefficient of kinetic friction μ_k.

Solution

Solving for μ_k gives

$$\mu_k = f_k N = f_k w\cos 25^{\circ} = f_k mg\cos 25^{\circ}$$

Substituting known values on the right-hand side of the equation,

$$\mu_k = 45.0\text{ N}(62\text{ kg})(9.80\text{ m/s}^2)(0.906) = 0.082$$

Discussion

This result is a little smaller than the coefficient listed in Table 5.1 for waxed wood on snow, but it is still reasonable since values of the coefficients of friction can vary greatly. In situations like this, where an object of mass m slides down a slope that makes an angle ϑ with the horizontal, friction is given by $f_k=\mu_k mg\cos\theta$. All objects will slide down a slope with constant acceleration under these circumstances. Proof of this is left for this chapter's Problems and Exercises.

Submicroscopic Explanations of Friction

The simpler aspects of friction dealt with so far are its macroscopic (large-scale) characteristics. Great strides have been made in the atomic-scale explanation of friction during the past several decades. Researchers are finding that the atomic nature of friction seems to have several fundamental characteristics. These characteristics not only explain some of the simpler aspects of friction—they also hold the potential for the development of nearly friction-free environments that could save hundreds of billions of dollars in energy which is currently being converted (unnecessarily) to heat.

Figure 5.3 illustrates one macroscopic characteristic of friction that is explained by microscopic (small-scale) research. We have noted that friction is proportional to the normal force, but not to the area in contact, a somewhat counterintuitive notion. When two rough surfaces are in contact, the actual contact area is a tiny fraction of the total area since only high spots touch. When a greater normal force is exerted, the actual contact area increases, and it is found that the friction is proportional to this area.

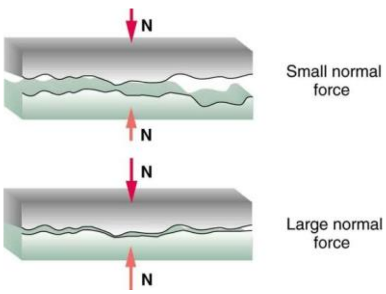


Figure **5.3:** Two rough surfaces in contact have a much smaller area of actual contact than their total area. When there is a greater normal force as a result of a greater applied force, the area of actual contact increases as does friction.

But the atomic-scale view promises to explain far more than the simpler features of friction. The mechanism for how heat is generated is now being determined. In other words, why do surfaces get warmer when rubbed? Essentially, atoms are linked with one another to form lattices. When surfaces rub, the surface atoms adhere and cause atomic lattices to vibrate—essentially creating sound waves that penetrate the material. The sound waves diminish with distance and their energy is converted into heat. Chemical reactions that are related to frictional wear can also occur between atoms and molecules on the surfaces. Figure 5.4 shows how the tip of a probe drawn across another material is deformed by atomic-scale friction. The force needed to drag the tip can be measured and is found to be related to shear stress, which will be discussed later in this chapter. The variation in shear stress is remarkable (more than a factor of 10^{12}) and difficult to predict theoretically, but shear stress is yielding a fundamental understanding of a large-scale phenomenon known since ancient times—friction.

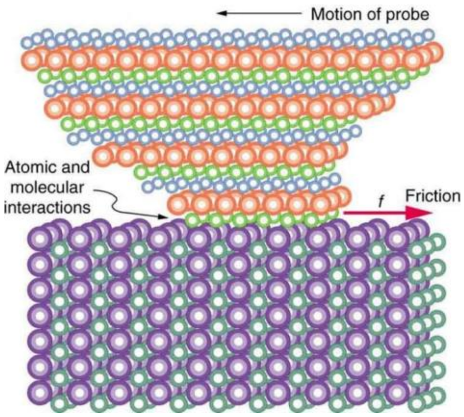


Figure **5.4:** The tip of a probe is deformed sideways by frictional force as the probe is dragged across a surface. Measurements of how the force varies for different materials are yielding fundamental insights into the atomic nature of friction.

5.3 Drag Force

Drag force F_D is found to be proportional to the square of the speed of the object. Mathematically

$$F_D \propto v^2$$

$$F_D = \frac{1}{2} C \rho A v^2$$

where C is the drag coefficient, A is the area of the object facing the fluid, and ρ is the density of the fluid.

The value of the drag coefficient, C is determined empirically, usually with the use of a wind tunnel.

The drag coefficient can depend upon velocity, but we will assume that it is a constant here. Table 5.2 lists some typical drag coefficients for a variety of objects. Notice that the drag coefficient is a dimensionless quantity. At highway speeds, over 50% of the power of a car is used to overcome air drag. The most fuel-efficient cruising speed is about 70–80 km/h (about 45–50 mi/h). For this reason, during the 1970s oil crisis in the United States, maximum speeds on highways were set at about 90 km/h (55 mi/h).

Object	C
Airfoil	0.05

Object	C
Toyota Camry	0.28
Ford Focus	0.32
Honda Civic	0.36
Ferrari Testarossa	0.37
Dodge Ram pickup	0.43
Sphere	0.45
Hummer H2 SUV	0.64
Skydiver (feet first)	0.70
Bicycle	0.90
Skydiver (horizontal)	1.0
Circular flat plate	1.12

Table **5.2:** Drag Coefficient Values Typical values of drag coefficient *C*.

Substantial research is under way in the sporting world to minimize drag. The dimples on golf balls are being redesigned as are the clothes that athletes wear. Bicycle racers and some swimmers and runners wear full bodysuits. Australian Cathy Freeman wore a full body suit in the 2000 Sydney Olympics, and won the gold medal for the 400 m race. Many swimmers in the 2008 Beijing Olympics wore (Speedo) body suits; it might have made a difference in breaking many world records. Most elite swimmers (and cyclists) shave their body hair. Such innovations can have the effect of slicing away milliseconds in a race, sometimes making the difference between a gold and a silver medal. One consequence is that careful and precise guidelines must be continuously developed to maintain the integrity of the sport.

Some interesting situations connected to Newton's second law occur when considering the effects of drag forces upon a moving object. For instance, consider a skydiver falling through air under the influence of gravity. The two forces acting on him are the force of gravity and the drag force (ignoring the buoyant force). The downward force of gravity remains constant regardless of the velocity at which the person is moving. However, as the person's velocity increases, the magnitude of the drag force increases until the magnitude of the drag force is equal to the gravitational force, thus producing a net force of zero. A zero net force means that there is no acceleration, as given by Newton's second law. At this point, the person's velocity remains constant and we say that the person has reached his *terminal velocity* (vt). Since F_D is proportional to the speed, a heavier skydiver must go faster for F_D to equal his weight. Let's see how this works out more quantitatively.

At the terminal velocity,

$$F_{net} = mg - F_D = ma = 0$$

Thus,

$$Mg = F_D.$$

Using the equation for drag force, we have

$$Mg = ½ \rho C A v^2$$

Solving for the velocity, we obtain

$$V = \surd(2mg)/ (\rho CA)$$

Assume the density of air is ρ=1.21 kg/m^3. A 75-kg skydiver descending headfirst will have an area approximately A =0.18m^2 and a drag coefficient of approximately C=0.70. We find that

$$v = \surd\ [2(75\ \text{kg})(9.80\ \text{m/s}^2)]/ [(1.21\ \text{kg/m}^3)(0.70)(0.18\text{m}^2)]$$

$$= 98\ \text{m/s}$$

$$= 350\ \text{km/h}$$

This means a skydiver with a mass of 75 kg achieves a maximum terminal velocity of about 350 km/h while traveling in a headfirst position, minimizing the area and his drag. In a spread-eagle position, that terminal velocity may decrease to about 200 km/h as the area increases. This terminal velocity becomes much smaller after the parachute opens.

Example - A Terminal Velocity

Find the terminal velocity of an 85-kg skydiver falling in a spread-eagle position.

Strategy

At terminal velocity, F_{net}=0. Thus the drag force on the skydiver must equal the force of gravity (the person's weight). Using the equation of drag force, we find

$$mg = 1/2\ \rho C A v^2$$

Therefore, the terminal velocity v_t can be written as

$$v_t = \surd(2mg) / (\rho CA)$$

Solution

All quantities are known except the person's projected area. This is an adult (85 kg) falling spread eagle. We can estimate the frontal area as

$$A=(2\ \text{m})(0.35\ \text{m})=0.70\text{m}^2$$

Using our equation for v_t, we find that,

$$v_t = \sqrt{[2(85kg)(9.80m/s^2)] / [(1.21kg/m^3)(1.0)(0.70m^2)]}$$

$$44\ m/s$$

Discussion

This result is consistent with the value for v_t mentioned earlier. The 75-kg skydiver going feet first had a v=98 m/s. He weighed less but had a smaller frontal area and so a smaller drag due to the air.

The size of the object that is falling through air presents another interesting application of air drag. If you fall from a 5-m high branch of a tree, you will likely get hurt—possibly fracturing a bone. However, a small squirrel does this all the time, without getting hurt. You don't reach a terminal velocity in such a short distance, but the squirrel does.

The above quadratic dependence of air drag upon velocity does not hold if the object is very small, is going very slow, or is in a denser medium than air. Then we find that the drag force is proportional just to the velocity. This relationship is given by Stokes' law, which states that

$$F_s=6\pi r\eta v$$

where r is the radius of the object, η is the viscosity of the fluid, and v is the object's velocity.

Good examples of this law are provided by microorganisms, pollen, and dust particles. Because each of these objects is so small, we find that many of these objects travel unaided only at a constant (terminal) velocity. Terminal velocities for bacteria (size about 1 μm) can be about 2 μm/s. To move at a greater speed, many bacteria swim using flagella (organelles shaped like little tails) that are powered by little motors embedded in the cell. Sediment in a lake can move at a greater terminal velocity (about 5 μm/s), so it can take days to reach the bottom of the lake after being deposited on the surface.

5.4 Elasticity: Stress and Strain

We now move from consideration of forces that affect the motion of an object (such as friction and drag) to those that affect an object's shape. If a bulldozer pushes a car into a wall, the car will not move but it will noticeably change shape. A change in shape due to the application of a force is a deformation. Even very small forces are known to cause some deformation. For small deformations, two important characteristics are observed. First, the object returns to its original shape when the force is removed—that is, the deformation is elastic for small deformations. Second, the size of the deformation is proportional to the force—that is, for small deformations, Hooke's law is obeyed. In equation form, Hooke's law is given by,

$$F=k\Delta L$$

where ΔL is the amount of deformation (the change in length, for example) produced by the force F, and k is a proportionality constant that depends on the shape and composition of the object and the direction of the force. Note that this force is a function of the deformation ΔL - it is not constant as a kinetic friction force is. Rearranging this to

$$\Delta L = F/k$$

makes it clear that the deformation is proportional to the applied force. Figure 5.5 shows the Hooke's law relationship between the extension ΔL of a spring or of a human bone. For metals or springs, the straight-line region in which Hooke's law pertains is much larger. Bones are brittle and the elastic region is small and the fracture abrupt. Eventually a large enough stress to the material will cause it to break or fracture. Tensile strength is the breaking stress that will cause permanent deformation or fracture of a material.

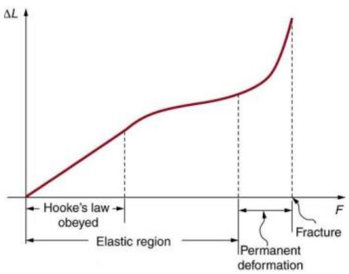


Figure 5.5: A graph of deformation ΔL versus applied force F. The straight segment is the linear region where Hooke's law is obeyed. The slope of the straight region is $1k$. For larger forces, the graph is curved but the deformation is still elastic— ΔL will return to zero if the force is removed. Still greater forces permanently deform the object until it finally fractures. The shape of the curve near fracture depends on several factors, including how the force F is applied. Note that in this graph the slope increases just before fracture, indicating that a small increase in F is producing a large increase in L near the fracture.

The proportionality constant k depends upon a number of factors for the material. For example, a guitar string made of nylon stretches when it is tightened, and the elongation ΔL is proportional to the force applied (at least for small deformations). Thicker nylon strings and ones made of steel stretch less for the same applied force, implying they have a larger k (see Figure 5.6). Finally, all three strings return to their normal lengths when the force is removed, provided the deformation is small. Most materials will behave in this manner if the deformation is less than about 0.1% or about 1 part in 10^3.

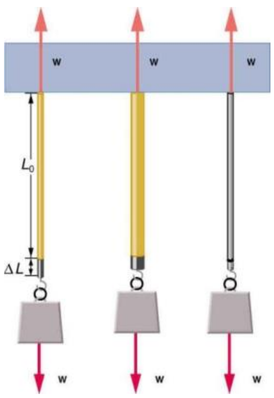


Figure 5.6: The same force, in this case a weight (w), applied to three different guitar strings of identical length produces the three different deformations shown as shaded segments. The string on the left is thin nylon, the one in the middle is thicker nylon, and the one on the right is steel.

We now consider three specific types of deformations: changes in length (tension and compression), sideways shear (stress), and changes in volume. All deformations are assumed to be small unless otherwise stated.

Changes in Length—Tension and Compression: Elastic Modulus

A change in length ΔL is produced when a force is applied to a wire or rod parallel to its length L_0, either stretching it (a tension) or compressing it. (See Figure 5.7.)

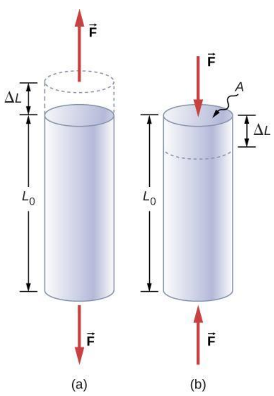


Figure 5.7: (a) Tension. The rod is stretched a length ΔL when a force is applied parallel to its length. (b) Compression. The same rod is compressed by forces with the same magnitude in the opposite direction. For very small deformations and uniform materials, ΔL is approximately the same for the same magnitude of tension or compression. For larger deformations, the cross-sectional area changes as the rod is compressed or stretched.

Experiments have shown that the change in length (ΔL) depends on only a few variables. As already noted, ΔL is proportional to the force F and depends on the substance from which the object is made. Additionally, the change in length is proportional to the original length $L0$ and inversely proportional to the cross-sectional area of the wire or rod. For example, a long guitar string will stretch more than a short one, and a thick string will stretch less than a thin one. We can combine all these factors into one equation for ΔL:

$$\Delta L = (1/Y)\ (F/A) L_0$$

where ΔL is the change in length, F the applied force, Y is a factor, called the elastic modulus or Young's modulus, that depends on the substance, A is the cross-sectional area, and $L0$ is the original length. Table 5.3 lists values of Y for several materials—those with a large Y are said to have a large tensile stiffness because they deform less for a given tension or compression.

Material	**Young's modulus (tension–compression)*Y* (10^9N/m2)**	**Shear modulus *S*(10^9N/m2)**	**Bulk modulus *B* (10^9N/m2)**
Aluminum	70	25	75
Bone – tension	16	80	8
Bone – compression	9		
Brass	90	35	75
Brick	15		
Concrete	20		
Glass	70	20	30
Granite	45	20	45
Hair (human)	10		
Hardwood	15	10	
Iron, cast	100	40	90
Lead	16	5	50
Marble	60	20	70
Nylon	5		
Polystyrene	3		
Silk	6		
Spider thread	3		

Material	Young's modulus (tension–compression)Y (10^9N/m2)	Shear modulus S(10^9N/m2)	Bulk modulus B (10^9N/m2)
Steel	210	80	130
Tendon	1		
Acetone			0.7
Ethanol			0.9
Glycerin			4.5
Mercury			25
Water			2.2

Table **5.3:** Elastic Moduli

Young's moduli are not listed for liquids and gases in Table 5.3 because they cannot be stretched or compressed in only one direction. Note that there is an assumption that the object does not accelerate, so that there are actually two applied forces of magnitude *F* acting in opposite directions. For example, the strings in Figure 5.7 are being pulled down by a force of magnitude *w* and held up by the ceiling, which also exerts a force of magnitude *w*.

Example - The Stretch of a Long Cable

Suspension cables are used to carry gondolas at ski resorts. (See Figure 5.8) Consider a suspension cable that includes an unsupported span of 3020 m. Calculate the amount of stretch in the steel cable. Assume that the cable has a diameter of 5.6 cm and the maximum tension it can withstand is 3.0×10^6N.

Figure **5.8:** Gondolas travel along suspension cables at the Gala Yuzawa ski resort in Japan. (credit: Rudy Herman, Flickr)

Strategy

The force is equal to the maximum tension, or $F = 3.0\times10^6$N. The cross-sectional area is,

$$\pi r2 = 2.46\times10^{-3}\text{m}^2$$

The equation $\Delta L = (1/Y)\ (F/A)L_0$ can be used to find the change in length.

Solution

All quantities are known. Thus,

$$\Delta L = (1/\ (210\times10^9\ \text{N/m}^2)\ )\ (3.0\times10^6\text{N} / 2.46\times10^{-3}\text{m}^2)\ (3020\ \text{m})$$

$$= 18\ \text{m}.$$

Discussion

This is quite a stretch, but only about 0.6% of the unsupported length. Effects of temperature upon length might be important in these environments.

Example - Calculating Deformation: How Much Does Your Leg Shorten When You Stand on It?

Calculate the change in length of the upper leg bone (the femur) when a 70.0 kg man supports 62.0 kg of his mass on it, assuming the bone to be equivalent to a uniform rod that is 40.0 cm long and 2.00 cm in radius.

Strategy

The force is equal to the weight supported, or

$$F = mg = (62.0\ \text{kg})(9.80\ \text{m/s}^2) = 607.6\ \text{N}$$

and the cross-sectional area is $\pi r^2 = 1.257\times10^{-3}\ \text{m}^2$. The equation $\Delta L = (1/Y)\ (F/A)L_0$ can be used to find the change in length.

Solution

All quantities except ΔL are known. Note that the compression value for Young's modulus for bone must be used here. Thus,

$$\Delta L = (1/\ 9\times10^9\text{N/m}^2)\ (607.6\ \text{N} / 1.257\times10^{-3}\text{m}^2)(0.400\ \text{m})$$

$$= 2\times10^{-5}\text{m}.$$

Discussion

This small change in length seems reasonable, consistent with our experience that bones are rigid. In fact, even the rather large forces encountered during strenuous physical activity do not compress or bend bones by large amounts. Although bone is rigid compared with fat or muscle, several of the substances listed in Table 5.3 have larger values of Young's modulus Y. In other words, they are more rigid.

The equation for change in length is traditionally rearranged and written in the following form:

$$F/A = Y\,(\Delta L/L_0)$$

The ratio of force to area, F/A, is defined as stress (measured in N/m^2), and the ratio of the change in length to length, $\Delta L/L_0$, is defined as strain (a unitless quantity). In other words,

$$\text{Stress} = Y \times \text{strain}$$

In this form, the equation is analogous to Hooke's law, with stress analogous to force and strain analogous to deformation. If we again rearrange this equation to the form

$$F = YA\,/\,(\Delta L\,/L_0)$$

we see that it is the same as Hooke's law with a proportionality constant

$$k = YA/L_0$$

This general idea—that force and the deformation it causes are proportional for small deformations—applies to changes in length, sideways bending, and changes in volume.

Sideways Stress: Shear Modulus

Figure 5.9 illustrates what is meant by a sideways stress or a *shearing force*. Here the deformation is called Δx and it is perpendicular to L_0, rather than parallel as with tension and compression. Shear deformation behaves similarly to tension and compression and can be described with similar equations. The expression for shear deformation is

$$\Delta x = (1/S)\,(F/A)\,L_0$$

where S is the shear modulus (see Table 5.3) and F is the force applied perpendicular to L_0 and parallel to the cross-sectional area A. Again, to keep the object from accelerating, there are actually two equal and opposite forces F applied across opposite faces, as illustrated in Figure 5.9. The equation is logical—for example, it is easier to bend a long thin pencil (small A) than a short thick one, and both are more easily bent than similar steel rods (large S).

Shear Deformation

$$\Delta x = (1/S)\,(F/A)L_0$$

where S is the shear modulus and F is the force applied perpendicular to L_0 and parallel to the cross-sectional area A.

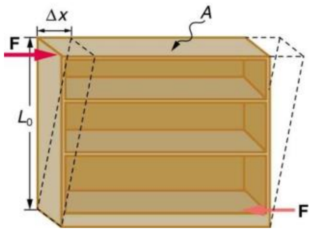


Figure 5.9: Shearing forces are applied perpendicular to the length L_0 and parallel to the area A, producing a deformation Δx. Vertical forces are not shown, but it should be kept in mind that in addition to the two shearing forces, **F**, there must be supporting forces to keep the object from rotating. The distorting effects of these supporting forces are ignored in this treatment. The weight of the object also is not shown, since it is usually negligible compared with forces large enough to cause significant deformations.

Examination of the shear moduli in Table 5.3 reveals some telling patterns. For example, shear moduli are less than Young's moduli for most materials. Bone is a remarkable exception. Its shear modulus is not only greater than its Young's modulus, but it is as large as that of steel. This is why bones are so rigid.

Example - Calculating Force Required to Deform: That Nail Does Not Bend Much Under a Load

Find the mass of the picture hanging from a steel nail as shown in Figure 5.10, given that the nail bends only 1.80 μm. (Assume the shear modulus is known to two significant figures.)

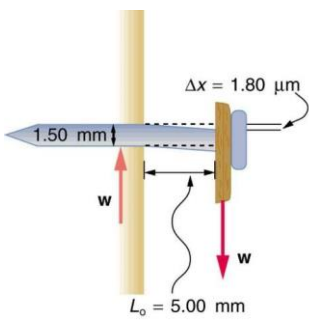


Figure 5.10: Side view of a nail with a picture hung from it. The nail flexes very slightly (shown much larger than actual) because of the shearing effect of the supported weight. Also shown is the upward force of the wall on the nail, illustrating that there are equal and opposite forces applied across opposite cross sections of the nail.

Strategy

The force F on the nail (neglecting the nail's own weight) is the weight of the picture w. If we can find w, then the mass of the picture is just wg. The equation $\Delta x = (1/S)\ (F/A)\ L_0$ can be solved for F.

Solution

Solving the equation $\Delta x = (1/S)\ (F/A)\ L_0$ for F, we see that all other quantities can be found:

$$F = (SA/L_0)\ \Delta x$$

S is found in Table 5.3 and is $S = 80 \times 10^9$ N/m . The radius r is 0.750 mm (as seen in the figure), so the cross-sectional area is,

$$A = \pi r^2 = 1.77\times10^{-6}\text{m}^2$$

The value for L_0 is also shown in the figure. Thus,

$$F = (80\times10^9\ \text{N/m}^2)\ (1.77\times10^{-6}\ \text{m}^2)\ (5.00\times10^{-3}\text{m})\ (1.80\times10^{-6}\text{m}) = 51\ \text{N}$$

This 51 N force is the weight w of the picture, so the picture's mass is

$$m = w/g = F/g = 5.2\ \text{kg}.$$

Discussion

This is a fairly massive picture, and it is impressive that the nail flexes only 1.80 μm—an amount undetectable to the unaided eye.

Changes in Volume: Bulk Modulus

An object will be compressed in all directions if inward forces are applied evenly on all its surfaces as in Figure 5.11. It is relatively easy to compress gases and extremely difficult to compress liquids and solids. For example, air in a wine bottle is compressed when it is corked. But if you try corking a brim-full bottle, you cannot compress the wine—some must be removed if the cork is to be inserted. The reason for these different compressibilities is that atoms and molecules are separated by large empty spaces in gases but packed close together in liquids and solids. To compress a gas, you must force its atoms and molecules closer together. To compress liquids and solids, you must actually compress their atoms and molecules, and very strong electromagnetic forces in them oppose this compression.

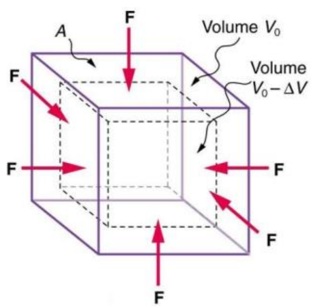


Figure **5.11:** An inward force on all surfaces compresses this cube. Its change in volume is proportional to the force per unit area and its original volume, and is related to the compressibility of the substance.

We can describe the compression or volume deformation of an object with an equation. First, we note that a force "applied evenly" is defined to have the same stress, or ratio of force to area F/A on all surfaces. The deformation produced is a change in volume ΔV, which is found to behave very similarly to the shear, tension, and compression previously discussed. (This is not surprising, since a compression of the entire object is equivalent to compressing each of its three dimensions.) The relationship of the change in volume to other physical quantities is given by

$$\Delta V = (1/B)\ (F/A) V_0$$

where B is the bulk modulus (see Table 5.3), V_0 is the original volume, and F/A is the force per unit area applied uniformly inward on all surfaces. Note that no bulk moduli are given for gases.

What are some examples of bulk compression of solids and liquids? One practical example is the manufacture of industrial-grade diamonds by compressing carbon with an extremely large force per unit area. The carbon atoms rearrange their crystalline structure into the more tightly packed pattern of diamonds. In nature, a similar process occurs deep underground, where extremely large forces result from the weight of overlying material. Another natural source of large compressive forces is the pressure created by the weight of water, especially in deep parts of the oceans. Water exerts an inward force on all surfaces of a submerged object, and even on the water itself. At great depths, water is measurably compressed, as the following example illustrates.

Example - Calculating Change in Volume with Deformation: How Much Is Water Compressed at Great Ocean Depths?

Calculate the fractional decrease in volume ($\Delta V/V_0$) for seawater at 5.00 km depth, where the force per unit area is $5.00\times10^7 \text{N/m}^2$.

Strategy

Equation $\Delta V = (1/B)\ (F/A) V_0$ is the correct physical relationship. All quantities in the equation except $\Delta V/V_0$ are known.

Solution

Solving for the unknown $\Delta V/V_0$ gives

$$\Delta V/V_0 = (1/B)\ (F/A)$$

Substituting known values with the value for the bulk modulus *B* from Table 5.3,

$$\Delta V/V_0 = (5.00\times10^{7}\text{N/m2}) / (2.2\times10^{9}\text{N/m2})$$

$$= 0.023 = 2.3\%.$$

Discussion

Although measurable, this is not a significant decrease in volume considering that the force per unit area is about 500 atmospheres (1 million pounds per square foot). Liquids and solids are extraordinarily difficult to compress.

Conversely, very large forces are created by liquids and solids when they try to expand but are constrained from doing so—which is equivalent to compressing them to less than their normal volume. This often occurs when a contained material warms up, since most materials expand when their temperature increases. If the materials are tightly constrained, they deform or break their container. Another very common example occurs when water freezes. Water, unlike most materials, expands when it freezes, and it can easily fracture a boulder, rupture a biological cell, or crack an engine block that gets in its way.

Other types of deformations, such as torsion or twisting, behave analogously to the tension, shear, and bulk deformations considered here.

For examples and answers, please refer to OpenStax.com questions and answers given on their website or to the College Physics 2e - https://openstax.org/details/books/college-physics-2e

Chapter 6

6.0 Objectives

At the end of this lesson, students should be able to,

- Calculate rotation angle and angular velocity.
- Calculate centripetal acceleration.
- Calculate centripetal force.
- Identify fictitious force and non-inertial frames.
- Apply Newton's Universal Law of Gravitation.
- Apply Kepler's laws.

6.1 Introduction

This chapter covers curved motion, uniform circular motion, motion in a circular path at constant speed and relevant calculations.

Pure *rotational motion* occurs when points in an object move in circular paths centered on one point. Pure *translational motion* is motion with no rotation. Some motion combines both types, such as a rotating hockey puck moving along ice.

In this chapter, we consider situations where the object does not land but moves in a curve. We begin the study of uniform circular motion by defining two angular quantities needed to describe rotational motion.

6.2 Rotation Angle

When objects rotate about some axis—for example, when the CD (compact disc) in Figure 6.1 rotates about its center—each point in the object follows a circular arc. Consider a line from the center of the CD to its edge. Each pit used to record sound along this line moves through the same angle in the same amount of time. The rotation angle is the amount of rotation and is analogous to linear distance. We define the rotation angle $\Delta\vartheta$ to be the ratio of the arc length to the radius of curvature:

$$\Delta\vartheta = \Delta sr.$$

Figure 6.1: All points on a CD travel in circular arcs. The pits along a line from the center to the edge all move through the same angle $\Delta\vartheta$ in a time Δt.

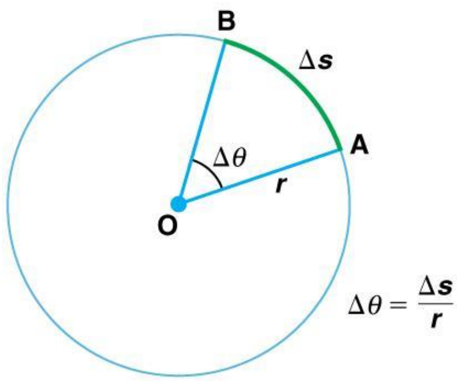


Figure 6.2: The radius of a circle is rotated through an angle $\Delta\vartheta$. The arc length Δs is described on the circumference.

The arc length Δs is the distance traveled along a circular path as shown in Figure 6.2. Note that r is the radius of curvature of the circular path.

We know that for one complete revolution, the arc length is the circumference of a circle of radius r. The circumference of a circle is $2\pi r$. Thus, for one complete revolution the rotation angle is

$$\Delta\vartheta = 2\pi r/r = 2\pi$$

This result is the basis for defining the units used to measure rotation angles, $\Delta\vartheta$ to be radians (rad), defined so that,

$$2\pi \text{ rad} = 1 \text{ revolution}$$

A comparison of some useful angles expressed in both degrees and radians is shown in Table 6.1:

Degree Measures	**Radian Measure**
30^{o}	$\pi 6$
60^{o}	$\pi 3$
90^{o}	$\pi 2$
120^{o}	$2\pi 3$
135^{o}	$3\pi 4$
180^{o}	π

Table 6.1: Comparison of Angular Units

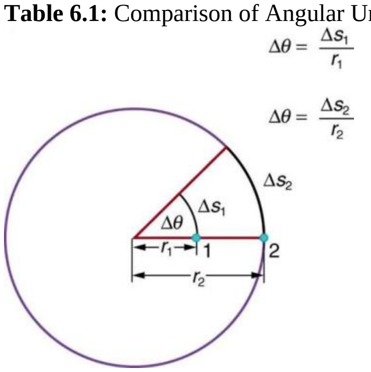


Figure 6.3: Points 1 and 2 rotate through the same angle ($\Delta\theta$), but point 2 moves through a greater arc length (Δs) because it is at a greater distance from the center of rotation (r).

If $\Delta\theta=2\pi$ rad, then the CD has made one complete revolution, and every point on the CD is back at its original position. Because there are 360º in a circle or one revolution, the relationship between radians and degrees is thus,

$$2\pi \text{ rad} = 360^{o}$$

so that

$$1\text{rad} = 360^{o} / 2\pi \approx 57.3^{o}$$

6.3 Angular Velocity

How fast is an object rotating? We define angular velocity ω as the rate of change of an angle. In symbols, this is

$$\omega = \Delta\vartheta/\Delta t$$

where an angular rotation $\Delta\vartheta$ takes place in a time Δt. The greater the rotation angle in a given amount of time, the greater the angular velocity. The units for angular velocity are radians per second (rad/s).

Angular velocity ω is analogous to linear velocity v. To get the precise relationship between angular and linear velocity, we again consider a pit on the rotating CD. This pit moves an arc length Δs in a time Δt, and so it has a linear velocity,

$$v = \Delta s/\Delta t$$

From $\Delta\vartheta = \Delta s/r$ we see that $\Delta s = r\Delta\vartheta$. Substituting this into the expression for v gives

$$v = (r\Delta\vartheta)/\Delta t = r\omega$$

We write this relationship in two different ways and gain two different insights:

$$v = r\omega \text{ or } \omega = v/r$$

The first relationship in $v = r\omega$ or $\omega = v/r$ states that the linear velocity v is proportional to the distance from the center of rotation, thus, it is largest for a point on the rim (largest r), as you might expect. We can also call this linear speed v of a point on the rim the *tangential speed*. The second relationship in $v = r\omega$ or $\omega = v/r$ can be illustrated by considering the tire of a moving car. Note that the speed of a point on the rim of the tire is the same as the speed v of the car. So the faster the car moves, the faster the tire spins—large v means a large ω, because $v = r\omega$. Similarly, a larger-radius tire rotating at the same angular velocity (ω) will produce a greater linear speed (v) for the car.

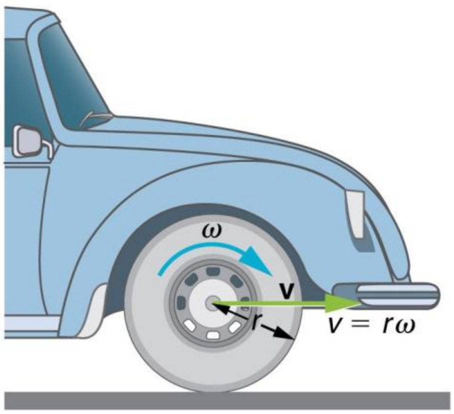


Figure 6.4: A car moving at a velocity v to the right has a tire rotating with an angular velocity ω.The speed of the tread of the tire relative to the axle is v, the same as if the car were jacked up. Thus, the car moves forward at linear velocity $v=r\omega$, where r is the tire radius. A larger angular velocity for the tire means a greater velocity for the car.

Example - How Fast Does a Car Tire Spin?

Calculate the angular velocity of a 0.300 m radius car tire when the car travels at 15.0m/s (about 54km/h). See Figure 6.4.

Strategy

Because the linear speed of the tire rim is the same as the speed of the car, we have v=15.0 m/s. The radius of the tire is given to be r=0.300 m. Knowing v and r, we can use the second relationship in $v=r\omega$, $\omega=v/r$ to calculate the angular velocity.

Solution

To calculate the angular velocity, we will use the following relationship:

$$\omega=vr.$$

Substituting the knowns,

$$\omega= (15.0\text{m/s}) /(0.300\text{m}) = 50.0\text{rad/s}.$$

Discussion

When we cancel units in the above calculation, we get 50.0/s. But the angular velocity must have units of rad/s. Because radians are actually unitless (radians are defined as a ratio of distance), we can simply insert them into the answer for the angular velocity. Also note that if an earth mover with much larger tires, say 1.20 m in radius, were moving at the same speed of 15.0 m/s, its tires would rotate more slowly. They would have an angular velocity

$$\omega=(15.0\text{m/s})/(1.20\text{m})=12.5 \text{ rad/s}.$$

Both ω and v have directions (hence they are angular and linear *velocities*, respectively). Angular velocity has only two directions with respect to the axis of rotation—it is either clockwise or counterclockwise. Linear velocity is tangent to the path, as illustrated in Figure 6.5.

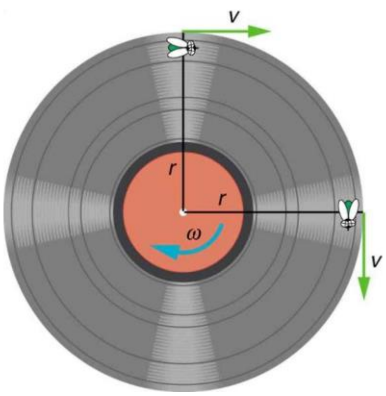


Figure 6.5: As an object moves in a circle, here a fly on the edge of an old-fashioned vinyl record, its instantaneous velocity is always tangent to the circle. The direction of the angular velocity is clockwise in this case.

6.4 Centripetal Acceleration

We know from kinematics that acceleration is a change in velocity, either in its magnitude or in its direction, or both. In uniform circular motion, the direction of the velocity changes constantly, so there is always an associated acceleration, even though the magnitude of the velocity might be constant. You experience this acceleration yourself when you turn a corner in your car. (If you hold the wheel steady during a turn and move at constant speed, you are in uniform circular motion.) What you notice is a sideways acceleration because you and the car are changing direction. The sharper the curve and the greater your speed, the more noticeable this acceleration will become. In this section we examine the direction and magnitude of that acceleration.

Figure 6.6 shows an object moving in a circular path at constant speed. The direction of the instantaneous velocity is shown at two points along the path. Acceleration is in the direction of the change in velocity, which points directly toward the center of rotation (the center of the circular path). This pointing is shown with the vector diagram in the figure. We call the acceleration of an object moving in uniform circular motion (resulting from a net external force) the centripetal acceleration(a_c); centripetal means "toward the center" or "center seeking."

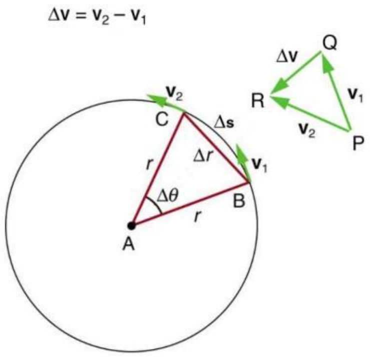


Figure 6.6: The directions of the velocity of an object at two different points are shown, and the change in velocity $\Delta \mathbf{v}$ is seen to point directly toward the center of curvature. (See small inset.) Because $\mathbf{a}_c = \Delta \mathbf{v}/\Delta t$, the acceleration is also toward the center; $\mathbf{a}_c$ is called centripetal acceleration. (Because $\Delta\vartheta$ is very small, the arc length Δs is equal to the chord length Δr for small time differences.)

The direction of centripetal acceleration is toward the center of curvature, but what is its magnitude? Note that the triangle formed by the velocity vectors and the one formed by the radii r and Δs are similar. Both the triangles ABC and PQR are isosceles triangles (two equal sides). The two equal sides of the velocity vector triangle are the speeds $v_1 = v_2 = v$. Using the properties of two similar triangles, we obtain

$$\Delta v/v = \Delta s/r.$$

Acceleration is $\Delta v/\Delta t$, and so we first solve this expression for Δv:

$$\Delta v = (v/r)\Delta s.$$

Then we divide this by Δt, yielding

$$\Delta v/\Delta t = (v/r) \times (\Delta s/\Delta t)$$

Finally, noting that $\Delta v/\Delta t = a_c$ and that $\Delta s/\Delta t = v$, the linear or tangential speed, we see that the magnitude of the centripetal acceleration is

$a_c = v^2/r$ acceleration is greater at high speeds and in sharp curves (smaller radius), as you have noticed when driving a car. But it is a bit surprising that a_c is proportional to speed squared, implying, for example, that it is four times as hard to take a curve at 100 km/h than at 50 km/h. A sharp corner has a small radius, so that a_c is greater for tighter turns, as you have probably noticed.

It is also useful to express a_c in terms of angular velocity. Substituting $v=r\omega$ into the above expression, we find $a_c=(r\omega)^2/r = r\omega^2$. We can express the magnitude of centripetal acceleration using either of two equations:

$$a_c = v^2/r \; ; \;\; a_c = r\omega^2$$

Recall that the direction of a_c is toward the center. You may use whichever expression is more convenient, as illustrated in examples below.

A centrifuge (see Figure 6.7 b) is a rotating device used to separate specimens of different densities. High centripetal acceleration significantly decreases the time it takes for separation to occur, and makes separation possible with small samples. Centrifuges are used in a variety of applications in science and medicine, including the separation of single cell suspensions such as bacteria, viruses, and blood cells from a liquid medium and the separation of macromolecules, such as DNA and protein, from a solution. Centrifuges are often rated in terms of their centripetal acceleration relative to acceleration due to gravity (g); maximum centripetal acceleration of several hundred thousand g is possible in a vacuum. Human centrifuges, extremely large centrifuges, have been used to test the tolerance of astronauts to the effects of accelerations larger than that of Earth’s gravity.

Example - How Does the Centripetal Acceleration of a Car Around a Curve Compare with That Due to Gravity?

What is the magnitude of the centripetal acceleration of a car following a curve of radius 500 m at a speed of 25.0 m/s (about 90 km/h)? Compare the acceleration with that due to gravity for this fairly gentle curve taken at highway speed. See Figure 6.7 (a).

Strategy

Because v and r are given, the first expression in $a_c = v^2/r$; $a_c = r\omega^2$ is the most convenient to use.

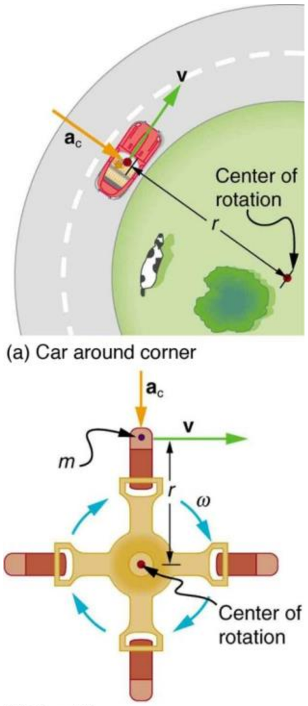
Solution

Entering the given values of v = 25.0m/s and r = 500 m into the first expression for a_c gives

$$a_c = v^2/r = (25.0\text{m/s})^2/\,(500\text{ m}) = 1.25\text{m/s}^2$$

Discussion

To compare this with the acceleration due to gravity (g = 9.80m/s^2), we take the ratio of a_c/g =(1.25m/s^2)/(9.80m/s^2) = 0.128. Thus, a_c = 0.128 g and is noticeable especially if you were not wearing a seat belt.



(b) Centrifuge

Figure 6.7: (a) The car following a circular path at constant speed is accelerated perpendicular to its velocity, as shown. The magnitude of this centripetal acceleration is found in the above example. (b) A particle of mass in a centrifuge is rotating at constant angular velocity . It must be accelerated perpendicular to its velocity, or it would continue in a straight line. The magnitude of the necessary acceleration is found in the example below:

Example - How Big Is the Centripetal Acceleration in an Ultracentrifuge?

Calculate the centripetal acceleration of a point 7.50 cm from the axis of an ultracentrifuge spinning at 7.5×10^4rev/min. Determine the ratio of this acceleration to that due to gravity. See Figure 6.7(b).

Strategy

The term rev/min stands for revolutions per minute. By converting this to radians per second, we obtain the angular velocity ω. Because r is given, we can use the second expression in the equation $a_c = v^2/r$; $a_c = r\omega^2$ to calculate the centripetal acceleration.

Solution

To convert 7.50×10^4 rev/min to radians per second, we use the facts that one revolution is 2πrad and one minute is 60.0 s. Thus,

$$\omega = 7.50\times10^4 \text{ rev/min} \times (2\pi \text{ rad/1 rev}) \times (1\text{min}/60.0 \text{ s}) = 7854 \text{ rad/s}.$$

Now the centripetal acceleration is given by the second expression in $a_c = v^2/r$; $a_c = r\omega^2$ as

$$a_c = r\omega^2$$

Converting 7.50 cm to meters and substituting known values gives

$$a_c = (0.0750 \text{ m})(7854 \text{ rad/s})^2 = 4.63\times10^6 \text{m/s}^2.$$

Note that the unitless radians are discarded in order to get the correct units for centripetal acceleration. Taking the ratio of a_c to g yields

$$a_c/g = 4.63\times10^6 / 9.80 = 4.72\times10^5$$

Discussion

This last result means that the centripetal acceleration is 472,000 times as strong as g. It is no wonder that such high ω centrifuges are called ultracentrifuges. The extremely large accelerations involved greatly decrease the time needed to cause the sedimentation of blood cells or other materials.

Of course, a net external force is needed to cause any acceleration, just as Newton proposed in his second law of motion. So a net external force is needed to cause a centripetal acceleration. In Centripetal Force, we will consider the forces involved in circular motion.

6.5 Centripetal Force

Any force or combination of forces can cause a centripetal or radial acceleration. Just a few examples are the tension in the rope on a tether ball, the force of Earth's gravity on the Moon, friction between roller skates and a rink floor, a banked roadway's force on a car, and forces on the tube of a spinning centrifuge.

Any net force causing uniform circular motion is called a centripetal force. The direction of a centripetal force is toward the center of curvature, the same as the direction of centripetal acceleration. According to Newton's second law of motion, net force is mass times acceleration:

net F=*ma*. For uniform circular motion, the acceleration is the centripetal acceleration — $a = ac$. Thus, the magnitude of centripetal force Fc is,

$$F_c = ma_c$$

By using the expressions for centripetal acceleration a_c from $a_c = v^2/r$; $a_c = r\omega^2$, we get two expressions for the centripetal force F_c in terms of mass, velocity, angular velocity, and radius of curvature:

$$Fc = m\,(v^2/r)$$
$$Fc = mr\omega^2$$

You may use whichever expression for centripetal force is more convenient. Centripetal force F_c is always perpendicular to the path and pointing to the center of curvature, because $\mathbf{a}_c$ is perpendicular to the velocity and pointing to the center of curvature.

Note that if you solve the first expression for *r*, you get,

$$r = (mv^2)/F_c$$

This implies that for a given mass and velocity, a large centripetal force causes a small radius of curvature—that is, a tight curve.

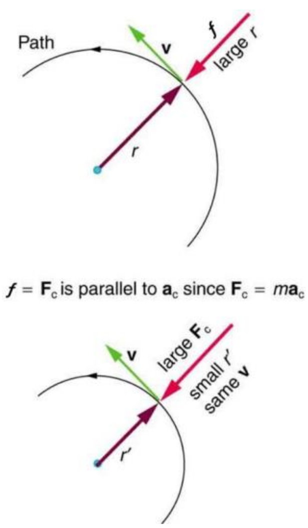


Figure 6.8: The frictional force supplies the centripetal force and is numerically equal to it. Centripetal force is perpendicular to velocity and causes uniform circular motion. The larger the F_c, the smaller the radius of curvature *r* and the sharper the curve. The second curve has the same *v*, but a larger Fc produces a smaller *r'*.

Example - What Coefficient of Friction Do Car Tires Need on a Flat Curve?

(a) Calculate the centripetal force exerted on a 900 kg car that negotiates a 500 m radius curve at 25.0 m/s.

(b) Assuming an unbanked curve, find the minimum static coefficient of friction, between the tires and the road, static friction being the reason that keeps the car from slipping (see Figure 6.9).

Strategy and Solution for (a)

We know that $F_c = (mv^2)/r$. Thus,

$$F_c = (mv^2)/r = (900 \text{ kg})(25.0 \text{ m/s})^2(500 \text{ m}) = 1125 \text{ N}.$$

Strategy for (b)

Figure 6.9 shows the forces acting on the car on an unbanked (level ground) curve. Friction is to the left, keeping the car from slipping, and because it is the only horizontal force acting on the car, the friction is the centripetal force in this case. We know that the maximum static friction (at which the tires roll but do not slip) is $\mu_s N$, where μ_s is the static coefficient of friction and N is the normal force. The normal force equals the car's weight on level ground, so that $N=mg$. Thus the centripetal force in this situation is,

$$F_c = f = \mu_s N = \mu_s mg$$

Now we have a relationship between centripetal force and the coefficient of friction. Using the first expression for F_c from the equation

$$F_c = m\,(v^2/r)$$

$$F_c = mr\omega^2$$

$$m\,(v^2/r) = \mu_s mg$$

We solve this for μ_s, noting that mass cancels, and obtain,

$$\mu_s = v^2/rg$$

Solution for (b)

Substituting the knowns,

$$\mu_s = (25.0 \text{ m/s})^2/\ [(500 \text{ m})(9.80 \text{ m/s}^2)] = 0.13.$$

(Because coefficients of friction are approximate, the answer is given to only two digits.)

Discussion

We could also solve part (a) using the first expression in $F_c = m\ (v2/r)$ *or* $F_c = mr\omega^2$ because *m*, *v*, and *r* are given. The coefficient of friction found in part (b) is much smaller than is typically found between tires and roads. The car will still negotiate the curve if the coefficient is greater than 0.13, because static friction is a responsive force, being able to assume a value less than but no more than $\mu_s N$. A higher coefficient would also allow the car to negotiate the curve at a higher speed, but if the coefficient of friction is less, the safe speed would be less than 25 m/s. Note that mass cancels, implying that in this example, it does not matter how heavily loaded the car is to negotiate the turn. Mass cancels because friction is assumed proportional to the normal force, which in turn is proportional to mass. If the surface of the road were banked, the normal force would be less as will be discussed below.

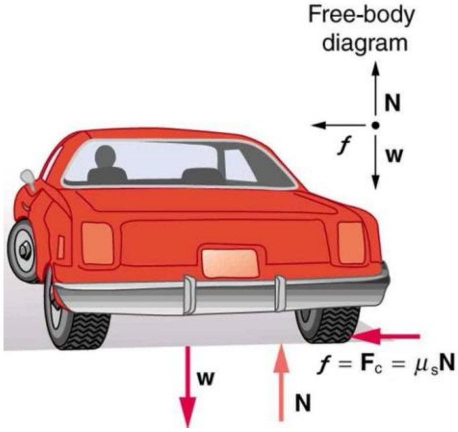


Figure 6.9: This car on level ground is moving away and turning to the left. The centripetal force causing the car to turn in a circular path is due to friction between the tires and the road. A minimum coefficient of friction is needed, or the car will move in a larger-radius curve and leave the roadway.

Let us now consider banked curves, where the slope of the road helps you negotiate the curve. See Figure 6.10. The greater the angle θ, the faster you can take the curve. Racetracks for bikes as well as cars, for example, often have steeply banked curves. In an “ideally banked curve,” the angle θ is such that you can negotiate the curve at a certain speed without the aid of friction between the tires and the road. We will derive an expression for θ for an ideally banked curve and consider an example related to it.

For ideal banking, the net external force equals the horizontal centripetal force in the absence of friction. The components of the normal force N in the horizontal and vertical directions must equal the centripetal force and the weight of the car, respectively. In cases in which forces are not parallel, it is most convenient to consider components along perpendicular axes—in this case, the vertical and horizontal directions.

Figure 6.10 shows a free body diagram for a car on a frictionless banked curve. If the angle θ is ideal for the speed and radius, then the net external force will equal the necessary centripetal force. The only two external forces acting on the car are its weight **w** and the normal force of the road **N**. (A frictionless surface can only exert a force perpendicular to the surface—that is, a normal force.) These two forces must add to give a net external force that is horizontal toward the center of curvature and has magnitude mv^2/r. Because this is the crucial force and it is horizontal, we use a coordinate system with vertical and horizontal axes. Only the normal force has a horizontal component, and so this must equal the centripetal force—that is,

$$N \sin \theta = (mv^2)/r$$

Because the car does not leave the surface of the road, the net vertical force must be zero, meaning that the vertical components of the two external forces must be equal in magnitude and opposite in direction. From the figure, we see that the vertical component of the normal force is $N\cos\theta$, and the only other vertical force is the car's weight. These must be equal in magnitude; thus,

$$N \cos \theta = mg$$

Now we can combine the last two equations to eliminate N and get an expression for θ, as desired. Solving the second equation for $N = mg / (\cos \theta)$, and substituting this into the first yields

$$mg\ (\sin\theta/\cos\theta) = (mv^2)/r$$

$$mg \tan(\theta) = (mv^2)/r$$

$$\tan\theta = v^2/rg$$

Taking the inverse tangent gives

$$\theta = \tan^{-1}(v^2/rg) \quad \text{(ideally banked curve, no friction).}$$

This expression can be understood by considering how θ depends on v and r. A large θ will be obtained for a large v and a small r. That is, roads must be steeply banked for high speeds and sharp curves. Friction helps, because it allows you to take the curve at greater or lower speed than if the curve is frictionless. Note that θ does not depend on the mass of the vehicle.

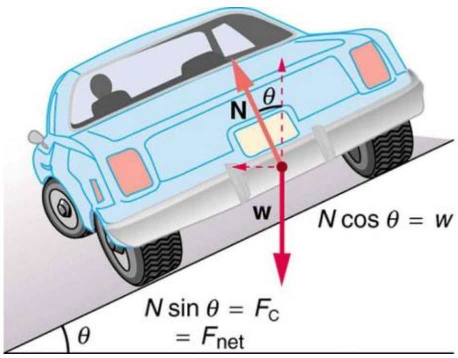


Figure 6.10: The car on this banked curve is moving away and turning to the left.

Example - What Is the Ideal Speed to Take a Steeply Banked Tight Curve?

Curves on some test tracks and racecourses, such as the Daytona International Speedway in Florida, are very steeply banked. This banking, with the aid of tire friction and very stable car configurations, allows the curves to be taken at very high speed. To illustrate, calculate the speed at which a 100 m radius curve banked at 65.0° should be driven if the road is frictionless.

Strategy

We first note that all terms in the expression for the ideal angle of a banked curve except for speed are known; thus, we need only rearrange it so that speed appears on the left-hand side and then substitute known quantities.

Solution

Starting with

$$\tan\theta = v^2/rg$$

we get

$$v = (\, r\, g \tan\theta)^{1/2}$$

Noting that tan 65.0º = 2.14, we obtain

$$v = [(100\text{ m})(9.80\text{ m/s}^2)(2.14)]^{1/2}$$

$$= 45.8\text{ m/s.}$$

Discussion

This is just about 165 km/h, consistent with a very steeply banked and rather sharp curve. Tire friction enables a vehicle to take the curve at significantly higher speeds.

6.6 Inertial Frame/Non-inertial Frame of Reference

What do, when taking off in a jet airplane, turning a corner in a car, riding a merry-go-round, and the circular motion of a tropical cyclone have in common? Each exhibits fictitious forces—unreal forces that arise from motion and may *seem* real, because the observer's frame of reference is accelerating or rotating.

When taking off in a jet, most people would agree it feels as if you are being pushed back into the seat as the airplane accelerates down the runway. Yet a physicist would say that *you* tend to remain stationary while the *seat* pushes forward on you, and there is no real force backward on you. An even more common experience occurs when you make a tight curve in your car—say, to the right. You feel as if you are thrown (that is, *forced*) toward the left relative to the car. Again, a physicist would say that *you* are going in a straight line, but the *car* moves to the right, and there is no real force on you to the left. Recall Newton's first law.

We can reconcile these points of view by examining the frames of reference used. Let us concentrate on people in a car (Figure 6.11). Passengers instinctively use the car as a frame of reference, while a physicist uses Earth. The physicist chooses Earth because it is very nearly an inertial frame of reference—one in which all forces are real (that is, in which all forces have an identifiable physical origin). In such a frame of reference, Newton's laws of motion take the form . The car is a non-inertial frame of reference because it is accelerated to the side. The force to the left sensed by car passengers is a fictitious force having no physical origin. There is nothing real pushing them left—the car, as well as the driver, is actually accelerating to the right.

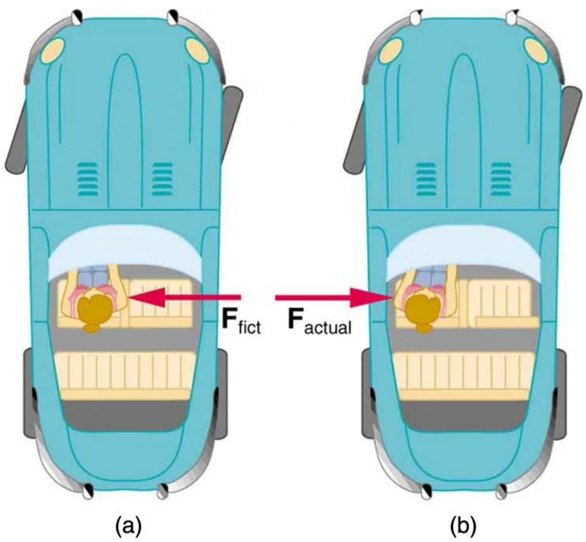


Figure 6.11: (a) The car driver feels herself forced to the left relative to the car when she makes a right turn. This is a fictitious force arising from the use of the car as a frame of reference. (b) In the Earth's frame of reference, the driver moves in a straight line, obeying Newton's first law, and the car moves to the right. There is no real force to the left on the driver relative to Earth. There is a real force to the right on the car to make it turn.

Let us now take a mental ride on a merry-go-round—specifically, a rapidly rotating playground merry-go-round (Figure 6.12). You take the merry-go-round to be your frame of reference because you rotate together. In that non-inertial frame, you feel a fictitious force, named centrifugal force **(**not to be confused with centripetal force**)**, trying to throw you off. You must hang on tightly to counteract the centrifugal force. In Earth's frame of reference, there is no force trying to throw you off. Rather you must hang on to make yourself go in a circle because otherwise you would go in a straight line, right off the merry-go-round.

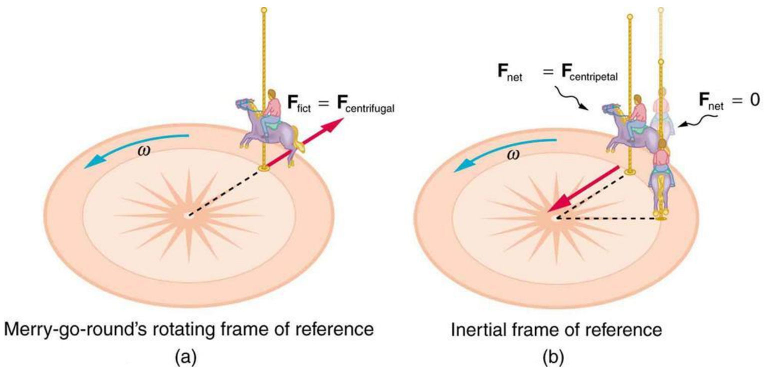


Figure 6.12: (a) A rider on a merry-go-round feels as if he is being thrown off. This fictitious force is called the centrifugal force—it explains the rider's motion in the rotating frame of reference. (b) In an inertial frame of reference and according to Newton's laws, it is his inertia that carries him off and not a real force (the unshaded rider has *F*net=0 and heads in a straight line). A real force, *F*centripetal, is needed to cause a circular path.

This inertial effect, carrying you away from the center of rotation if there is no centripetal force to cause circular motion, is put to good use in centrifuges (see Figure 6.13). A centrifuge spins a sample very rapidly, as mentioned earlier in this chapter. Viewed from the rotating frame of reference, the fictitious centrifugal force throws particles outward, hastening their sedimentation. The greater the angular velocity, the greater the centrifugal force. But what really happens is that the inertia of the particles carries them along a line tangent to the circle while the test tube is forced in a circular path by a centripetal force.

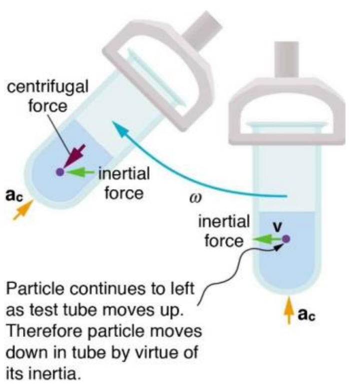


Figure 6.13: Centrifuges use inertia to perform their task. Particles in the fluid sediment come out because their inertia carries them away from the center of rotation. The large angular velocity of the centrifuge quickens the sedimentation. Ultimately, the particles will come into contact with the test tube walls, which will then supply the centripetal force needed to make them move in a circle of constant radius.

Let us now consider what happens if something moves in a frame of reference that rotates. For example, what if you slide a ball directly away from the center of the merry-go-round, as shown in Figure 6.14? The ball follows a straight path relative to Earth (assuming negligible friction) and a path curved to the right on the merry-go-round's surface. A person standing next to the merry-go-round sees the ball moving straight and the merry-go-round rotating underneath it. In the merry-go-round's frame of reference, we explain the apparent curve to the right by using a fictitious force, called the Coriolis force, that causes the ball to curve to the right. The fictitious Coriolis force can be used by anyone in that frame of reference to explain why objects follow curved paths and allows us to apply Newton's Laws in non-inertial frames of reference.

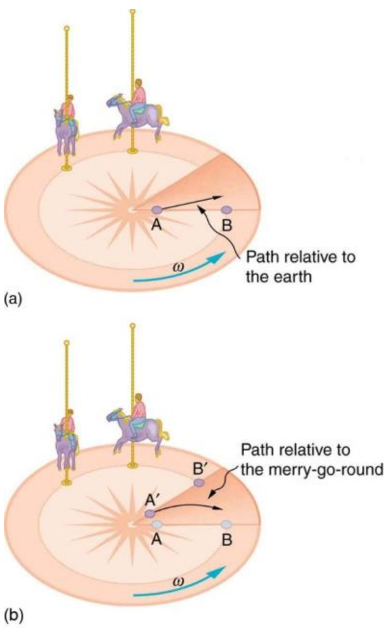


Figure 6.14: Looking down on the counterclockwise rotation of a merry-go-round, we see that a ball slid straight toward the edge follows a path curved to the right. The person slides the ball toward point B, starting at point A. Both points rotate to the shaded positions (A' and B') shown in the time that the ball follows the curved path in the rotating frame and a straight path in Earth's frame.

Up until now, we have considered Earth to be an inertial frame of reference with little or no worry about effects due to its rotation. Yet such effects *do* exist—in the rotation of weather systems, for example. Most consequences of Earth's rotation can be qualitatively understood by analogy with the merry-go-round. Viewed from above the North Pole, Earth rotates counterclockwise, as does the merry-go-round in Figure 6.14. As on the merry-go-round, any motion in Earth's northern hemisphere experiences a Coriolis force to the right. Just the opposite occurs in the southern hemisphere; there, the force is to the left. Because Earth's angular velocity is small, the Coriolis force is usually negligible, but for large-scale motions, such as wind patterns, it has substantial effects.

The Coriolis force causes hurricanes in the northern hemisphere to rotate in the counterclockwise direction, while the tropical cyclones (what hurricanes are called below the equator) in the southern hemisphere rotate in the clockwise direction. The terms hurricane, typhoon, and tropical storm are regionally specific names for tropical cyclones, storm systems characterized by low pressure centers, strong winds, and heavy rains. Figure 6.15 helps show how these rotations take place. Air flows toward any region of low pressure, and tropical cyclones contain particularly low pressures. Thus, winds flow toward the center of a tropical cyclone or a low-pressure weather system at the surface. In the northern hemisphere, these inward winds are deflected to the right, as shown in the figure, producing counterclockwise circulation at the surface for low-pressure zones of any type. Low pressure at the surface is associated with rising air, which also produces cooling and cloud formation, making low-pressure patterns quite visible from space. Conversely, wind circulation around high-pressure zones is clockwise in the northern hemisphere but is less visible because high pressure is associated with sinking air, producing clear skies.

The rotation of tropical cyclones and the path of a ball on a merry-go-round can just as well be explained by inertia and the rotation of the system underneath. When non-inertial frames are used, fictitious forces, such as the Coriolis force, must be invented to explain the curved path. There is no identifiable physical source for these fictitious forces. In an inertial frame, inertia explains the path, and no force is found to be without an identifiable source. Either view allows us to describe nature, but a view in an inertial frame is the simplest and truest, in the sense that all forces have real origins and explanations.

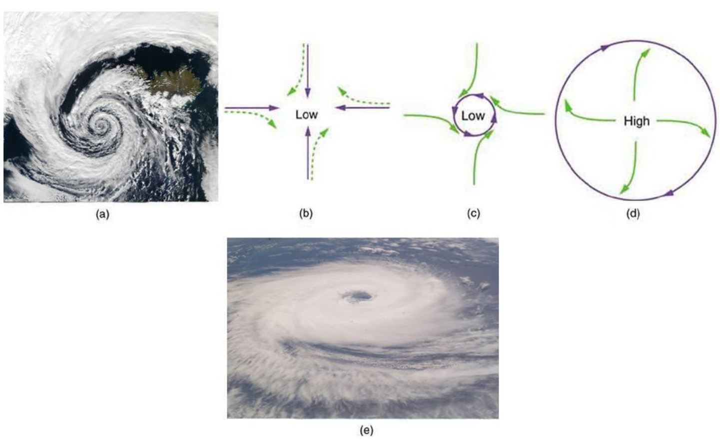


Figure 6.15: (a) The counterclockwise rotation of this northern hemisphere hurricane is a major consequence of the Coriolis force. (credit: NASA) (b) Without the Coriolis force, air would flow straight into a low-pressure zone, such as that found in tropical cyclones. (c) The Coriolis force deflects the winds to the right, producing a counterclockwise rotation. (d) Wind flowing away from a high-pressure zone is also deflected to the right, producing a clockwise rotation. (e) The

opposite direction of rotation is produced by the Coriolis force in the southern hemisphere, leading to tropical cyclones. (credit: NASA)

6.7 Newton's Universal Gravitational Law

What do aching feet, a falling apple, and the orbit of the Moon have in common? Each is caused by the gravitational force. Our feet are strained by supporting our weight—the force of Earth's gravity on us. An apple falls from a tree because of the same force acting a few meters above Earth's surface. And the Moon orbits Earth because gravity is able to supply the necessary centripetal force at a distance of hundreds of millions of meters. In fact, the same force causes planets to orbit the Sun, stars to orbit the center of the galaxy, and galaxies to cluster together. Gravity is another example of underlying simplicity in nature. It is the weakest of the four basic forces found in nature, and in some ways the least understood. It is a force that acts at a distance, without physical contact, and is expressed by a formula that is valid everywhere in the universe, for masses and distances that vary from the tiny to the immense.

The gravitational force is relatively simple. It is always attractive, and it depends only on the masses involved and the distance between them. Stated in modern language, Newton's universal law of gravitation states that every particle in the universe attracts every other particle with a force along a line joining them. The force is directly proportional to the product of their masses and inversely proportional to the square of the distance between them.

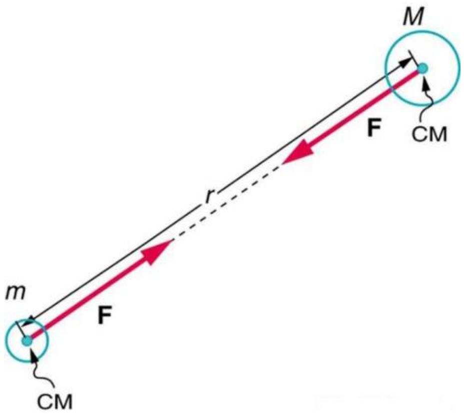


Figure 6.16: Gravitational attraction is along a line joining the centers of mass of these two bodies. The magnitude of the force is the same on each, consistent with Newton's third law.

The bodies we are dealing with tend to be large. To simplify the situation we assume that the body acts as if its entire mass is concentrated at one specific point called the center of mass (CM). For two bodies having masses *m* and *M* with a distance *r* between their centers of mass, the equation for Newton's universal law of gravitation is

$$F = G\,[(mM)/\,r^2]$$

where F is the magnitude of the gravitational force and G is a proportionality factor called the gravitational constant. G is a universal gravitational constant—that is, it is thought to be the same everywhere in the universe. It has been measured experimentally to be

$$G = 6.674\times10^{-11}\ \mathrm{N\cdot m^2/kg^2}$$

in SI units. Note that the units of G are such that a force in newtons is obtained from,

$$F = G\ [(mM)/\ r^2]$$

when considering masses in kilograms and distance in meters. For example, two 1.000 kg masses separated by 1.000 m will experience a gravitational attraction of 6.674×10^{-11}N. This is an extraordinarily small force. The small magnitude of the gravitational force is consistent with everyday experience. We are unaware that even large objects like mountains exert gravitational forces on us. In fact, our body weight is the force of attraction of the *entire Earth* on us with a mass of 6×10^{24}kg.

Recall that the acceleration due to gravity g is about 9.80 m/s^2 on Earth. We can now determine why this is so. The weight of an object mg is the gravitational force between it and Earth. Substituting mg for F in Newton's universal law of gravitation gives

$$Mg = G\ [(mM)/r^2]$$

where m is the mass of the object, M is the mass of Earth, and r is the distance to the center of Earth (the distance between the centers of mass of the object and Earth). See Figure 6.17. The mass m of the object cancels, leaving an equation for g:

$$g = G(M/r^2)$$

Substituting known values for Earth's mass and radius (to three significant figures),

$$g = (6.67\times10^{-11}\mathrm{N\cdot m^2kg^2}) \times [(5.98\times10^{24}\mathrm{kg})/\ (6.38\times10^6\mathrm{m})^2]$$

and we obtain a value for the acceleration of a falling body:

$$g = 9.80\mathrm{m/s^2}$$

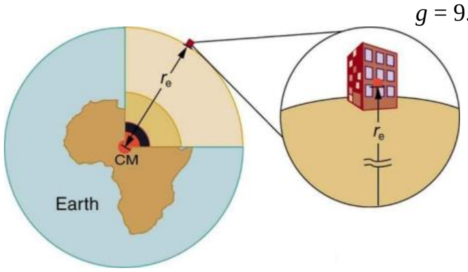


Figure 6.17: The distance between the centers of mass of Earth and an object on its surface is very nearly the same as the radius of Earth, because Earth is so much larger than the object.

This is the expected value *and is independent of the body's mass*. Newton's law of gravitation takes Galileo's observation that all masses fall with the same acceleration a step further, explaining the observation in terms of a force that causes objects to fall—in fact, in terms of a universally existing force of attraction between masses.

Why does Earth not remain stationary as the Moon orbits it? This is because, as expected from Newton's third law, if Earth exerts a force on the Moon, then the Moon should exert an equal and opposite force on Earth (see Figure 6.18). We do not sense the Moon's effect on Earth's motion, because the Moon's gravity moves our bodies right along with Earth but there are other signs on Earth that clearly show the effect of the Moon's gravitational force.

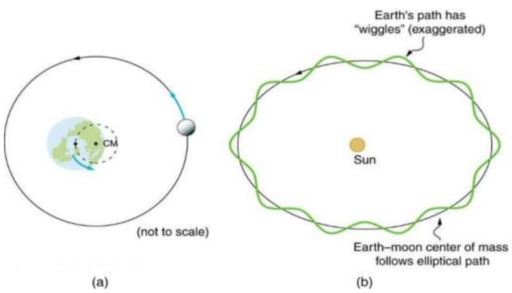


Figure 6.18: (a) Earth and the Moon rotate approximately once a month around their common center of mass. (b) Their center of mass orbits the Sun in an elliptical orbit, but Earth's path around the Sun has "wiggles" in it. Similar wiggles in the paths of stars have been observed and are considered direct evidence of planets orbiting those stars. This is important because the planets' reflected light is often too dim to be observed.

Tides

Ocean tides are one very observable result of the Moon's gravity acting on Earth. Figure 6.19 is a simplified drawing of the Moon's position relative to the tides. Because water easily flows on Earth's surface, a high tide is created on the side of Earth nearest to the Moon, where the Moon's gravitational pull is strongest. Why is there also a high tide on the opposite side of Earth? The answer is that Earth is pulled toward the Moon more than the water on the far side, because Earth is closer to the Moon. So, the water on the side of Earth closest to the Moon is pulled away from Earth, and Earth is pulled away from water on the far side. As Earth rotates, the tidal bulge (an effect of the tidal forces between an orbiting natural satellite and the primary planet that it orbits) keeps its orientation with the Moon. Thus, there are two tides per day (the actual tidal period is about 12 hours and 25.2 minutes), because the Moon moves in its orbit each day as well).

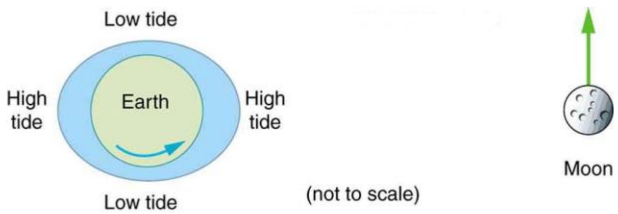


Figure 6.19: The Moon causes ocean tides by attracting the water on the near side more than Earth, and by attracting Earth more than the water on the far side. The distances and sizes are not to scale. For this simplified representation of the Earth-Moon system, there are two high and two low tides per day at any location, because Earth rotates under the tidal bulge.

The Sun also affects tides, although it has about half the effect of the Moon. However, the largest tides, called spring tides, occur when Earth, the Moon, and the Sun are aligned. The smallest tides, called neap tides, occur when the Sun is at a 90° angle to the Earth-Moon alignment.

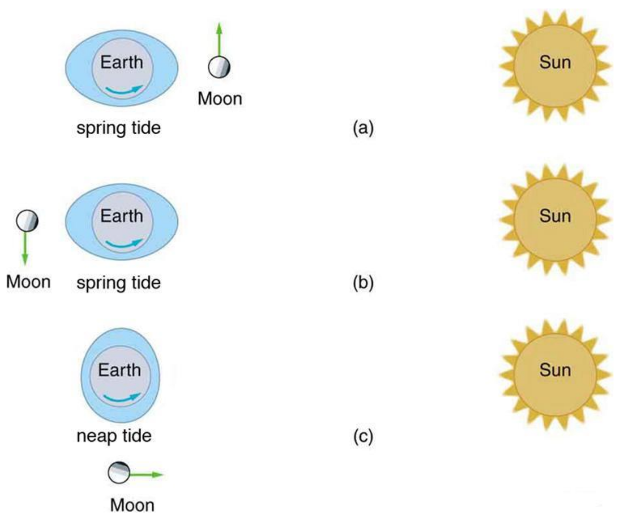


Figure 6.20: (a, b) Spring tides: The highest tides occur when Earth, the Moon, and the Sun are aligned. (c) Neap tide: The lowest tides occur when the Sun lies at 90° to the Earth-Moon alignment. Note that this figure is not drawn to scale.

Tides are not unique to Earth but occur in many astronomical systems. The most extreme tides occur where the gravitational force is the strongest and varies most rapidly, such as near black holes (see Figure 6.21). A few likely candidates for black holes have been observed in our galaxy. These have masses greater than the Sun but have diameters only a few kilometers across. The tidal forces near them are so great that they can actually tear matter from a companion star.

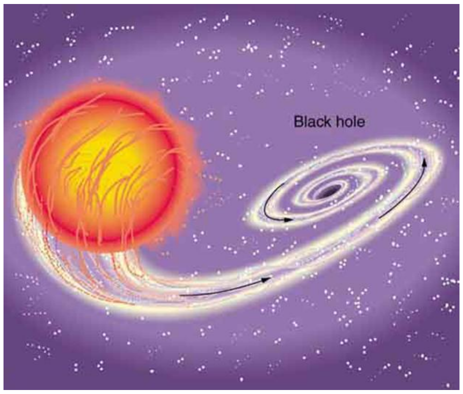


Figure 6.21: A black hole is an object with such strong gravity that not even light can escape it. This black hole was created by the supernova of one star in a two-star system. The tidal forces created by the black hole are so great that it tears matter from the companion star. This matter is compressed and heated as it is sucked into the black hole, creating light and X-rays observable from Earth.

“Weightlessness” and Microgravity

In contrast to the tremendous gravitational force near black holes is the apparent gravitational field experienced by astronauts orbiting Earth. What is the effect of “weightlessness” upon an astronaut who is in orbit for months? Or what about the effect of weightlessness upon plant growth? Weightlessness doesn’t mean that an astronaut is not being acted upon by the gravitational force. There is no “zero gravity” in an astronaut’s orbit. The term just means that the astronaut is in free-fall, accelerating with the acceleration due to gravity. If an elevator cable breaks, the passengers inside will be in free fall and will experience weightlessness. You can experience short periods of weightlessness in some rides in amusement parks.

Microgravity refers to an environment in which the apparent net acceleration of a body is small compared with that produced by Earth at its surface. Many interesting biology and physics topics have been studied over the past three decades in the presence of microgravity. Of immediate concern is the effect on astronauts of extended times in outer space, such as at the International Space Station. Researchers have observed that muscles will atrophy (waste away) in this environment. There is also a corresponding loss of bone mass. Study continues on cardiovascular adaptation to space flight. On Earth, blood pressure is usually higher in the feet than in the head, because the higher column of blood exerts a downward force on it, due to gravity. When standing, 70% of your blood is below the level of the heart, while in a horizontal position, just the opposite occurs. What difference does the absence of this pressure differential have upon the heart?

Some findings in human physiology in space can be clinically important to the management of diseases back on Earth. On a somewhat negative note, spaceflight is known to affect the human immune system, possibly making the crew members more vulnerable to infectious diseases. Experiments flown in space also have shown that some bacteria grow faster in microgravity than they do on Earth. However, on a positive note, studies indicate that microbial antibiotic production can increase by a factor of two in space-grown cultures. One hopes to be able to understand these mechanisms so that similar successes can be achieved on the ground. In another area of physics space research, inorganic crystals and protein crystals have been grown in outer space that have much higher quality than any grown-on Earth, so crystallography studies on their structure can yield much better results.

Plants have evolved with the stimulus of gravity and with gravity sensors. Roots grow downward and shoots grow upward. Plants might be able to provide a life support system for long duration space missions by regenerating the atmosphere, purifying water, and producing food. Some studies have indicated that plant growth and development are not affected by gravity, but there is still uncertainty about structural changes in plants grown in a microgravity environment.

The Cavendish Experiment: Then and Now

As previously noted, the universal gravitational constant G is determined experimentally. This definition was first done accurately by Henry Cavendish (1731–1810), an English scientist, in 1798, more than 100 years after Newton published his universal law of gravitation. The measurement of G is very basic and important because it determines the strength of one of the four forces in nature. Cavendish's experiment was very difficult because he measured the tiny gravitational attraction between two ordinary-sized masses (tens of kilograms at most), using apparatus like that in Figure 6.22. Remarkably, his value for G differs by less than 1% from the best modern value.

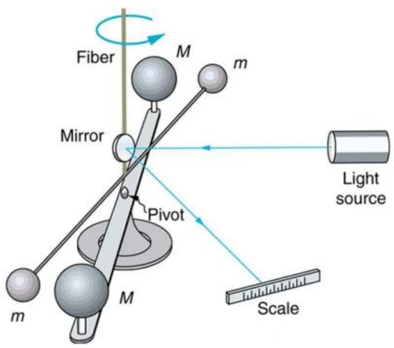


Figure 6.22: Cavendish used an apparatus like this to measure the gravitational attraction between the two suspended spheres (m) and the two on the stand (M) by observing the amount of torsion (twisting) created in the fiber. Distance between the masses can be varied to check the dependence of the force on distance. Modern experiments of this type continue to explore gravity.

One important consequence of knowing G was that an accurate value for Earth's mass could finally be obtained. This was done by measuring the acceleration due to gravity as accurately as possible and then calculating the mass of Earth M from the relationship Newton's universal law of gravitation gives

$$Mg = G\ [(mM)/r^2\ ]$$

where m is the mass of the object, M is the mass of Earth, and r is the distance to the center of Earth (the distance between the centers of mass of the object and Earth). The mass m of the object cancels, leaving an equation for g:

$$g=G[M/r^2]$$

Rearranging to solve for M yields

$$M= gr^2/G$$

So M can be calculated because all quantities on the right, including the radius of Earth r, are known from direct measurements.

The Cavendish experiment is also used to explore other aspects of gravity. One of the most interesting questions is whether the gravitational force depends on substance as well as mass—for example, whether one kilogram of lead exerts the same gravitational pull as one kilogram of water. A Hungarian scientist named Roland von Eötvös pioneered this inquiry early in the 20th century. He found, with an accuracy of five parts per billion, that the gravitational force does not depend on the substance. Such experiments continue today and have improved upon Eötvös' measurements. Cavendish-type experiments such as those of Eric Adelberger and others at the University of Washington, have also put severe limits on the possibility of a fifth force and have verified a major prediction of general relativity—that gravitational energy contributes to rest mass. Ongoing measurements there use a torsion balance and a parallel plate (not spheres, as Cavendish used) to examine how Newton's law of gravitation works over sub-millimeter distances. On this small-scale, do gravitational effects depart from the inverse square law? So far, no deviation has been observed.

Examples of gravitational orbits abound. Hundreds of artificial satellites orbit Earth together with thousands of pieces of debris. The Moon's orbit about Earth has intrigued humans from time immemorial. The orbits of planets, asteroids, meteors, and comets about the Sun are no less interesting. If we look further, we see almost unimaginable numbers of stars, galaxies, and other celestial objects orbiting one another and interacting through gravity.

All these motions are governed by gravitational force, and it is possible to describe them to various degrees of precision. Precise descriptions of complex systems must be made with large computers. However, we can describe an important class of orbits without the use of computers, and we shall find it instructive to study them. These orbits have the following characteristics:

1. *A small mass m orbits a much larger mass M*. This allows us to view the motion as if *M* were stationary—in fact, as if from an inertial frame of reference placed on *M* —without significant error. Mass *m* is the satellite of *M*, if the orbit is gravitationally bound.
2. *The system is isolated from other masses*. This allows us to neglect any small effects due to outside masses.

The conditions are satisfied, in good approximation, by Earth's satellites (including the Moon), by objects orbiting the Sun, and by the satellites of other planets. Historically, planets were studied first, and there is a classical set of three laws, called Kepler's laws of planetary motion, that describe the orbits of all bodies satisfying the two previous conditions (not just planets in our solar system). These descriptive laws are named for the German astronomer Johannes Kepler (1571–1630), who devised them after careful study (over some 20 years) of a large amount of meticulously recorded observations of planetary motion done by Tycho Brahe (1546–1601). Such careful collection and detailed recording of methods and data are hallmarks of good science. Data constitute the evidence from which new interpretations and meanings can be constructed.

6.8 Kepler's Laws of Planetary Motion

Kepler's First Law

The orbit of each planet about the Sun is an ellipse with the Sun at one focus.

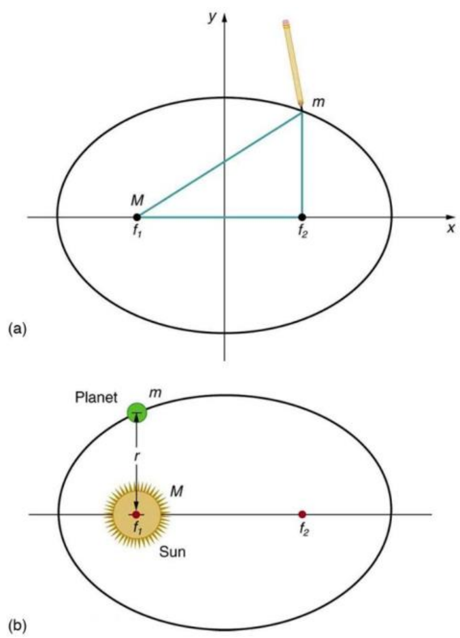


Figure 6.23: (a) An ellipse is a closed curve such that the sum of the distances from a point on the curve to the two foci (*f*1 and *f*2) is a constant. You can draw an ellipse as shown by putting a pin at each focus, and then placing a string around a pencil and the pins and tracing a line on paper. A circle is a special case of an ellipse in which the two foci coincide (thus any point on the circle is the same distance from the center). (b) For any closed gravitational orbit, *m* follows an elliptical path with *M* at one focus. Kepler's first law states this fact for planets orbiting the Sun.

Kepler's Second Law

Each planet moves so that an imaginary line drawn from the Sun to the planet sweeps out equal areas in equal times (see Figure 6.24).

Kepler's Third Law

The ratio of the squares of the periods of any two planets about the Sun is equal to the ratio of the cubes of their average distances from the Sun. In equation form, this is

$$T_1^2/T_2^2 = r_1^3/r_2^3$$

where T is the period (time for one orbit) and r is the average radius. This equation is valid only for comparing two small masses orbiting the same large one. Most importantly, this is a descriptive equation only, giving no information as to the cause of the equality.

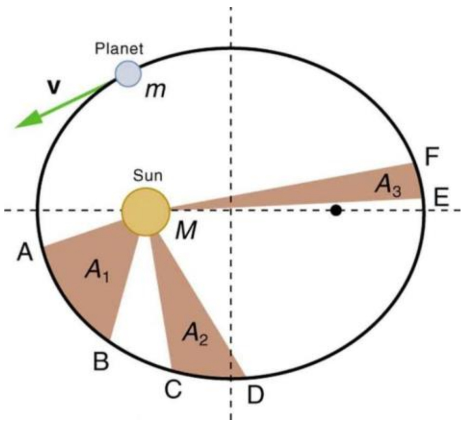


Figure 6.24: The shaded regions have equal areas. It takes equal times for m to go from A to B, from C to D, and from E to F. The mass m moves fastest when it is closest to M. Kepler's second law was originally devised for planets orbiting the Sun, but it has broader validity.

Note again that while, for historical reasons, Kepler's laws are stated for planets orbiting the Sun, they are actually valid for all bodies satisfying the two previously stated conditions.

Example - Find the Time for One Orbit of an Earth Satellite

Given that the Moon orbits Earth each 27.3 d and that it is an average distance of 3.84×10^8m from the center of Earth, calculate the period of an artificial satellite orbiting at an average altitude of 1500 km above Earth's surface.

Strategy

The period, or time for one orbit, is related to the radius of the orbit by Kepler's third law, given in mathematical form in $T_1^2/T_2^2 = r_1^3/r_2^3$. Let us use the subscript 1 for the Moon and the subscript 2 for the satellite. We are asked to find T_2. The given information tells us that the orbital radius of the Moon is $r_1=3.84\times10^8 \text{m}$, and that the period of the Moon is $T_1=27.3\text{ d}$. The height of the artificial satellite above Earth's surface is given, and so we must add the radius of Earth (6380 km) to get $r_2=(1500+6380)\text{km}=7880\text{km}$. Now all quantities are known, and so T_2 can be found.

Solution

Kepler's third law is

$$T_1^2/T_2^2 = r_1^3/r_2^3$$

To solve for T_2, we cross-multiply and take the square root yielding,

$$T_2^{\,2} = T_1^{\,2}/(r_2/r_1)^3$$

$$T_2 = T_1\,(r_2/r_1)^{3/2}$$

Substituting known values yields

$$T_2 = 27.3 \text{ d} \times 924.0 \text{ h/d}) \times (7880 \text{ km} / 3.84{\times}10^5 \text{km})^{3/2}$$

$$= 1.93 \text{ h}.$$

Discussion

This is a reasonable period for a satellite in a fairly low orbit. It is interesting that any satellite at this altitude will orbit in the same amount of time. This fact is related to the condition that the satellite's mass is small compared with that of Earth.

People immediately search for deeper meaning when broadly applicable laws, like Kepler's, are discovered. It was Newton who took the next giant step when he proposed the law of universal gravitation. While Kepler was able to discover *what* was happening, Newton discovered that gravitational force was the cause.

Derivation of Kepler's Third Law for Circular Orbits

We shall derive Kepler's third law, starting with Newton's laws of motion and his universal law of gravitation. The point is to demonstrate that the force of gravity is the cause for Kepler's laws (although we will only derive the third one).

Let us consider a circular orbit of a small mass m around a large mass M, satisfying the two conditions stated at the beginning of this section. Gravity supplies the centripetal force to mass m. Starting with Newton's second law applied to circular motion,

$$F_{\text{net}} = ma_c = mv^2/r.$$

The net external force on mass m is gravity, and so we substitute the force of gravity for Fnet:

$$G\,(mM/r^2) = m(v^2/r).$$

The mass m cancels, yield,

$$G(M/r) = v^2$$

The fact that m cancels out is another aspect of the oft-noted fact that at a given location all masses fall with the same acceleration. Here we see that at a given orbital radius r, all masses orbit at the same speed. (This was implied by the result of the preceding worked example.) Now, to get at Kepler's third law, we must get the period T into the equation. By definition, period T is the time for one complete orbit. Now the average speed v is the circumference divided by the period—that is,

$$v = (2\pi r)/T$$

Substituting this into the previous equation gives,

$$G\,(M/r) = 4\pi^2 r^2/T^2$$

Solving for T^2 yields

$$T^2 = (4\pi^2/GM)r^3$$

Using subscripts 1 and 2 to denote two different satellites, and taking the ratio of the last equation for satellite 1 to satellite 2 yields,

$$T_1^2/T_2^2 = r_1^3/r_2^3$$

This is Kepler's third law. Note that Kepler's third law is valid only for comparing satellites of the same parent body, because only then does the mass of the parent body M cancel.

Now consider what we get if we solve,

$T^2 = (4\pi^2/GM)r3$ for the ratio r^3/T^2. We obtain a relationship that can be used to determine the mass M of a parent body from the orbits of its satellites:

$$r^3/T^2 = (G/4\pi^2)\,M$$

If r and T are known for a satellite, then the mass M of the parent can be calculated. This principle has been used extensively to find the masses of heavenly bodies that have satellites. Furthermore, the ratio r^3/T^2 should be a constant for all satellites of the same parent body (because $r^3/T^2 = (G/4\pi^2)\,M$). (See Table 6.2).

It is clear from Table 6.2 that the ratio of r^3/T^2 is constant, at least to the third digit, for all listed satellites of the Sun, and for those of Jupiter. Small variations in that ratio have two causes—uncertainties in the r and T data, and perturbations of the orbits due to other bodies. Interestingly, those perturbations can be—and have been—used to predict the location of new planets and moons. This is another verification of Newton's universal law of gravitation.

Parent	**Satellite**	**Average orbital radius *r*(km)**	**Period *T(y)***	**r^3 / T^2 (km^3 / y^2)**
Earth	Moon	3.84×10$_5$	0.07481	1.01×10$_{19}$
Sun	Mercury	5.79×10$_7$	0.2409	3.34×10$_{24}$
	Venus	1.082×10$_8$	0.6150	3.35×10$_{24}$

Parent	Satellite	Average orbital radius *r*(km)	Period *T(y)*	r^3 / T^2 (km^3 / y^2)
	Earth	1.496×10^8	1.000	3.35×10^{24}
	Mars	2.279×10^8	1.881	3.35×10^{24}
	Jupiter	7.783×10^8	11.86	3.35×10^{24}
	Saturn	1.427×10^9	29.46	3.35×10^{24}
	Neptune	4.497×10^9	164.8	3.35×10^{24}
	Pluto	5.90×10^9	248.3	3.33×10^{24}
Jupiter	Io	4.22×10^5	0.00485 (1.77 d)	3.19×10^{21}
	Europa	6.71×10^5	0.00972 (3.55 d)	3.20×10^{21}
	Ganymede	1.07×10^6	0.0196 (7.16 d)	3.19×10^{21}
	Callisto	1.88×10^6	0.0457 (16.19 d)	3.20×10^{21}

Table 6.2: Orbital Data and Kepler's Third Law

The universal law of gravitation is a good example of a physical principle that is very broadly applicable. That single equation for the gravitational force describes all situations in which gravity acts. It gives a cause for a vast number of effects, such as the orbits of the planets and moons in the solar system. It epitomizes the underlying unity and simplicity of physics.

For examples and answers, please refer to OpenStax.com questions and answers given on their website or to the College Physics 2e - https://openstax.org/details/books/college-physics-2e

Chapter 8

8.0 Objectives

At the end of this lesson, students should be able to,

- Identify conservation of momentum.
- Identify elastic and inelastic collisions.
- Calculate momentum and impulse.
- Apply knowledge in momentum to calculate energy values and velocities in two dimensional collisions.

8.1 Introduction

This chapter covers the momentum, impulse, collisions and related calculations.

We use the term momentum in various ways in everyday language, and most of these ways are consistent with its precise scientific definition. We speak of sports teams or politicians gaining and maintaining the momentum to win. We also recognize that momentum has something to do with collisions. Generally, momentum implies a tendency to continue on course—to move in the same direction—and is associated with great mass and speed.

Momentum, like energy, is important because it is conserved. Only a few physical quantities are conserved in nature, and studying them yields fundamental insight into how nature works, as we shall see in our study of momentum.

8.2 Linear Momentum

The scientific definition of linear momentum is consistent with most people's intuitive understanding of momentum: a large, fast-moving object has greater momentum than a smaller, slower object. Linear momentum is defined as the product of a system's mass multiplied by its velocity. In symbols, linear momentum is expressed as

$$\mathbf{P} = m\mathbf{v}.$$

Momentum is directly proportional to the object's mass and also its velocity. Thus, the greater an object's mass or the greater its velocity, the greater its momentum. Momentum $\mathbf{p}$ is a vector having the same direction as the velocity $\mathbf{v}$. The SI unit for momentum is kg·m/s.

Example - Calculating Momentum: A Football Player and a Football

(a) Calculate the momentum of a 110-kg football player running at 8.00 m/s. (b) Compare the player's momentum with the momentum of a hard-thrown 0.410-kg football that has a speed of 25.0 m/s.

Strategy

No information is given regarding direction, and so we can calculate only the magnitude of the momentum, p. In both parts of this example, the magnitude of momentum can be calculated directly from the definition of momentum given in the equation, which becomes,

$$P = mv$$

when only magnitudes are considered.

Solution for (a)

To determine the momentum of the player, substitute the known values for the player's mass and speed into the equation.

$$p_{\text{player}} = (110\text{ kg})(8.00\text{ m/s}) = 880\text{ kg}\cdot\text{m/s}$$

Solution for (b)

To determine the momentum of the ball, substitute the known values for the ball's mass and speed into the equation.

$$p_{\text{ball}} = (0.410\text{ kg})(25.0\text{ m/s}) = 10.3\text{ kg}\cdot\text{m/s}$$

The ratio of the player's momentum to that of the ball is

$$p_{player} / p_{\text{ball}} = 880 / 10.3 = 85.9.$$

Discussion

Although the ball has greater velocity, the player has a much greater mass. Thus, the momentum of the player is much greater than the momentum of the football, as you might guess. As a result, the player's motion is only slightly affected if he catches the ball. We shall quantify what happens in such collisions in terms of momentum in later sections.

8.3 Momentum and Newton's Second Law

The importance of momentum, unlike the importance of energy, was recognized early in the development of classical physics. Momentum was deemed so important that it was called the "quantity of motion." Newton actually stated his second law of motion in terms of momentum: The net external force equals the change in momentum of a system divided by the time over which it changes. Using symbols, this law is

$$\mathbf{F}_{\text{net}} = \Delta\mathbf{p}\Delta t,$$

where $\mathbf{F}_{\text{net}}$ is the net external force, $\Delta\mathbf{p}$ is the change in momentum, and Δt is the change in time.

Force and momentum are intimately related. Force acting over time can change momentum, and Newton's second law of motion, can be stated in its most broadly applicable form in terms of momentum. Momentum continues to be a key concept in the study of atomic and subatomic particles in quantum mechanics.

This statement of Newton's second law of motion includes the more familiar $\mathbf{F}_{\text{net}} = m\mathbf{a}$ as a special case. We can derive this form as follows. First, note that the change in momentum $\Delta\mathbf{p}$ is given by

$$\Delta\mathbf{p} = \Delta(m\mathbf{v}).$$

If the mass of the system is constant, then

$$\Delta(m\mathbf{v})=m\Delta\mathbf{v}.$$

So that for constant mass, Newton's second law of motion becomes,

$$\mathbf{F}_{\text{net}} = \Delta\mathbf{p}\Delta t = m\Delta\mathbf{v}\Delta t.$$

Because $\Delta\mathbf{v}\Delta t = \mathbf{a}$, we get the familiar equation

$$\mathbf{F}_{\text{net}} = m\mathbf{a}$$

when the mass of the system is constant.

Newton's second law of motion stated in terms of momentum is more generally applicable because it can be applied to systems where the mass is changing, such as rockets, as well as to systems of constant mass.

8.4 Impulse: Change in Momentum

Change in momentum equals the average net external force multiplied by the time this force acts.

$$\Delta\mathbf{p} = \mathbf{F}_{\text{net}}\Delta t$$

The quantity $\mathbf{F}_{\text{net}}\Delta t$ is given the name impulse.

There are many ways in which an understanding of impulse can save lives, or at least limbs. The dashboard padding in a car, and certainly the airbags, allow the net force on the occupants in the car to act over a much longer time when there is a sudden stop. The momentum change is the same for an occupant, whether an air bag is deployed or not, but the force (to bring the occupant to a stop) will be much less if it acts over a larger time. Cars today have many plastic components. One advantage of plastics is their lighter weight, which results in better gas mileage. Another advantage is that a car will crumple in a collision, especially in the event of a head-on collision. A longer collision time means the force on the car will be less. Deaths during car races decreased dramatically when the rigid frames of racing cars were replaced with parts that could crumple or collapse in the event of an accident.

Bones in a body will fracture if the force on them is too large. If you jump onto the floor from a table, the force on your legs can be immense if you land stiff-legged on a hard surface. Rolling on the ground after jumping from the table, or landing with a parachute, extends the time over which the force (on you from the ground) acts.

Example - Calculating Magnitudes of Impulses: Two Billiard Balls Striking a Rigid Wall

Two identical billiard balls strike a rigid wall with the same speed and are reflected without any change of speed. The first ball strikes perpendicular to the wall. The second ball strikes the wall at an angle of 30º from the perpendicular and bounces off at an angle of 30º from perpendicular to the wall.

(a) Determine the direction of the force on the wall due to each ball.

(b) Calculate the ratio of the magnitudes of impulses on the two balls by the wall.

Strategy for (a)

In order to determine the force on the wall, consider the force on the ball due to the wall using Newton's second law and then apply Newton's third law to determine the direction. Assume the *x*-axis to be normal to the wall and to be positive in the initial direction of motion. Choose the *y*-axis to be along the wall in the plane of the second ball's motion. The momentum direction and the velocity direction are the same.

Solution for (a)

The first ball bounces directly into the wall and exerts a force on it in the $+x$ direction. Therefore, the wall exerts a force on the ball in the $-x$ direction. The second ball continues with the same momentum component in the *y* direction, but reverses its *x*-component of momentum, as seen by sketching a diagram of the angles involved and keeping in mind the proportionality between velocity and momentum.

These changes mean the change in momentum for both balls is in the $-x$ direction, so the force of the wall on each ball is along the $-x$ direction.

Strategy for (b)

Calculate the change in momentum for each ball, which is equal to the impulse imparted to the ball.

Solution for (b)

Let u be the speed of each ball before and after collision with the wall, and m the mass of each ball. Choose the x-axis and y-axis as previously described, and consider the change in momentum of the first ball which strikes perpendicular to the wall.

$$p_{xi} = mu;\ p_{yi} = 0$$

$$p_{xf} = -mu;\ p_{yf} = 0$$

Impulse is the change in momentum vector. Therefore, the x-component of impulse is equal to $-2mu$ and the y-component of impulse is equal to zero.

Now consider the change in momentum of the second ball.

$$p_{xi} = mu \cos 30^{\circ};\ p_{yi} = -mu \sin 30^{\circ}$$

$$p_{xf} = -mu \cos 30^{\circ};\ p_{yf} = -mu \sin 30^{\circ}$$

It should be noted here that while p_x changes sign after the collision, p_y does not. Therefore, the x-component of impulse is equal to $-2mu \cos 30^{\circ}$ and the y-component of impulse is equal to zero.

The ratio of the magnitudes of the impulse imparted to the balls is

$$2mu / (2mu \cos 30^{\circ}) = 2 / \sqrt{3} = 1.155.$$

Discussion

The direction of impulse and force is the same as in the case of (a); it is normal to the wall and along the negative x-direction. Making use of Newton's third law, the force on the wall due to each ball is normal to the wall along the positive x -direction.

Our definition of impulse includes an assumption that the force is constant over the time interval Δt. *Forces are usually not constant*. Forces vary considerably even during the brief time intervals considered. It is, however, possible to find an average effective force F_{eff} that produces the same result as the corresponding time-varying force. Figure 8.1 shows a graph of what an actual force looks like as a function of time for a ball bouncing off the floor. The area under the curve has units of momentum and is equal to the impulse or change in momentum between times t_1 and t_2. That area is equal to the area inside the rectangle bounded by F_{eff}, t_1, and t_2. Thus, the impulses and their effects are the same for both the actual and effective forces.

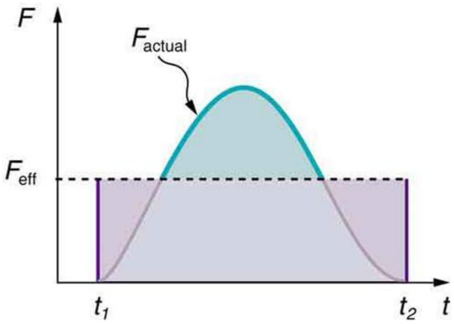


Figure 8.1: A graph of force versus time with time along the x-axis and force along the y-axis for an actual force and an equivalent effective force. The areas under the two curves are equal.

Momentum is an important quantity because it is conserved. Yet it was not conserved in the examples in Impulse and Linear Momentum and Force, where large changes in momentum were produced by forces acting on the system of interest. Under what circumstances is momentum conserved?

The answer to this question entails considering a sufficiently large system. It is always possible to find a larger system in which total momentum is constant, even if momentum changes for components of the system. If a football player runs into the goalpost in the end zone, there will be a force on him that causes him to bounce backward. The backward momentum felt by an object or person exerting force on another object is often called a recoil. However, the Earth also recoils—conserving momentum—because of the force applied to it through the goalpost. Because Earth is many orders of magnitude more massive than the player, its recoil is immeasurably small and can be neglected in any practical sense, but it is real nevertheless.

Consider what happens if the masses of two colliding objects are more similar than the masses of a football player and Earth—for example, one car bumping into another, as shown in Figure 8.2. Both cars are coasting in the same direction when the lead car (labeled m_2) is bumped by the trailing car (labeled m_1). The only unbalanced force on each car is the force of the collision. (Assume that the effects due to friction are negligible.) Car 1 slows down as a result of the collision, losing some momentum, while car 2 speeds up and gains some momentum. We shall now show that the total momentum of the two-car system remains constant.

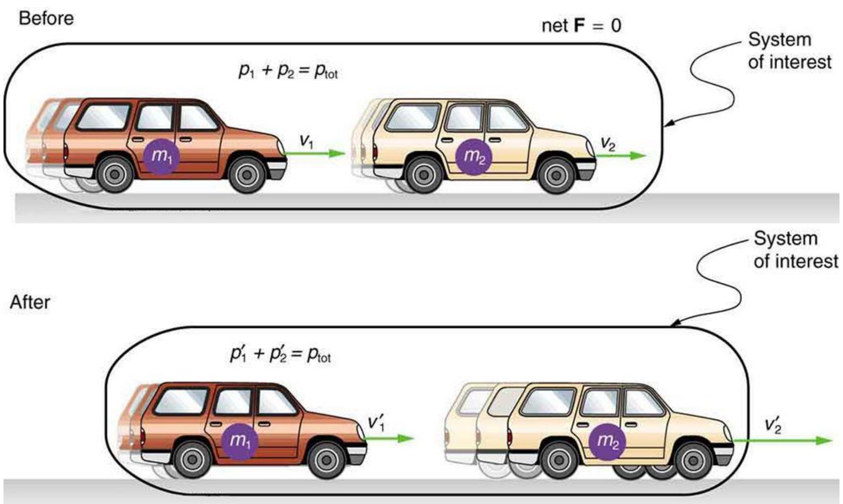


Figure 8.2: A car of mass m_1 moving with a velocity of v_1 bumps into another car of mass m_2 and velocity v_2 that it is following. As a result, the first car slows down to a velocity of v'_1 and the second speeds up to a velocity of v'_2. The momentum of each car is changed, but the total momentum p_{tot} of the two cars is the same before and after the collision (if you assume friction is negligible).

Using the definition of impulse, the change in momentum of car 1 is given by,

$$\Delta p_1 = F_1 \Delta t,$$

where F_1 is the force on car 1 due to car 2, and Δt is the time the force acts (the duration of the collision). Intuitively, it seems obvious that the collision time is the same for both cars, but it is only true for objects traveling at ordinary speeds. This assumption must be modified for objects travelling near the speed of light, without affecting the result that momentum is conserved.

Similarly, the change in momentum of car 2 is,

$$\Delta p_2 = F_2 \Delta t$$

where F_2 is the force on car 2 due to car 1, and we assume the duration of the collision Δt is the same for both cars. We know from Newton's third law that $F_2 = -F_1$, and so

$$\Delta p_2 = -F_1 \Delta t = -\Delta p_1$$

Thus, the changes in momentum are equal and opposite, and

$$\Delta p_1 + \Delta p_2 = 0$$

Because the changes in momentum add to zero, the total momentum of the two-car system is constant. That is,

$$p_1 + p_2 = \text{constant}$$

$$p_1 + p_2 = p'_1 + p'_2$$

where p'_1 and p'_2 are the momenta of cars 1 and 2 after the collision. (We often use primes to denote the final state.)

This result—that momentum is conserved—has validity far beyond the preceding one-dimensional case. It can be similarly shown that total momentum is conserved for any isolated system, with any number of objects in it. In equation form, the conservation of momentum principle for an isolated system is written,

$$\mathbf{p}_{\text{tot}} = \text{constant},$$

or

$$\mathbf{p}_{\text{tot}} = \mathbf{p}'_{\text{tot}} \text{ (Isolated System)}$$

where $\mathbf{p}_{\text{tot}}$ is the total momentum (the sum of the momenta of the individual objects in the system) and $\mathbf{p}'_{\text{tot}}$ is the total momentum later. (The total momentum can be shown to be the momentum of the center of mass of the system.) An isolated system is defined to be one for which the net external force is zero ($\mathbf{F}_{\text{net}} = 0$).

8.5 Isolated System

An isolated system is defined to be one for which the net external force is zero ($\mathbf{F}_{\text{net}}=0$). Perhaps an easier way to see that momentum is conserved for an isolated system is to consider Newton’s second law in terms of momentum, $\mathbf{F}_{\text{net}} = \Delta\mathbf{p}_{\text{tot}}\Delta t$. For an isolated system, ($\mathbf{F}_{\text{net}} = 0$); thus, $\Delta\mathbf{p}_{\text{tot}} = 0$, and $\mathbf{p}_{\text{tot}}$ is constant.

We have noted that the three length dimensions in nature—x, y, and z—are independent, and it is interesting to note that momentum can be conserved in different ways along each dimension. For example, during projectile motion and where air resistance is negligible, momentum is conserved in the horizontal direction because horizontal forces are zero and momentum is unchanged. But along the vertical direction, the net vertical force is not zero and the momentum of the projectile is not conserved. (See Figure 8.3.) However, if the momentum of the projectile-Earth system is considered in the vertical direction, we find that the total momentum is conserved.

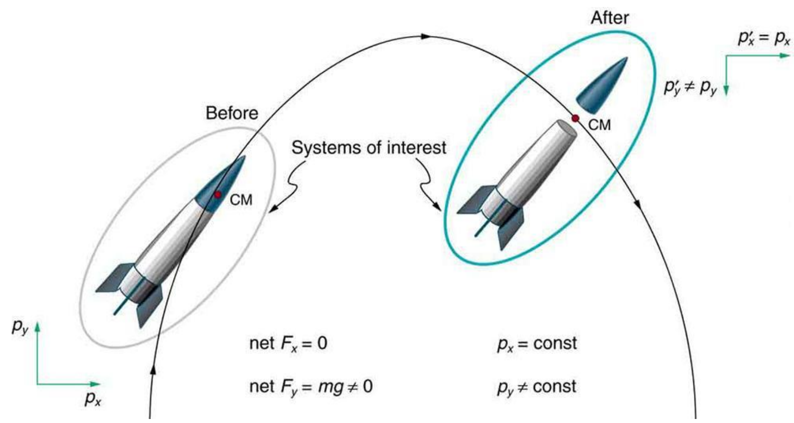


Figure 8.3: The horizontal component of a projectile's momentum is conserved if air resistance is negligible, even in this case where a space probe separates. The forces causing the separation are internal to the system, so that the net external horizontal force *Fx*–net is still zero. The vertical component of the momentum is not conserved, because the net vertical force *Fy*–net is not zero. In the vertical direction, the space probe-Earth system needs to be considered and we find that the total momentum is conserved. The center of mass of the space probe takes the same path it would if the separation did not occur.

The conservation of momentum principle can be applied to systems as different as a comet striking Earth and a gas containing huge numbers of atoms and molecules. Conservation of momentum is violated only when the net external force is not zero. But another larger system can always be considered in which momentum is conserved by simply including the source of the external force. For example, in the collision of two cars considered above, the two-car system conserves momentum while each one-car system does not.

Subatomic Collisions and Momentum

The conservation of momentum principle not only applies to macroscopic objects, it is also essential to our explorations of atomic and subatomic particles. Giant machines hurl subatomic particles at one another, and researchers evaluate the results by assuming conservation of momentum (among other things).

On the small scale, we find that particles and their properties are invisible to the naked eye but can be measured with our instruments, and models of these subatomic particles can be constructed to describe the results. Momentum is found to be a property of all subatomic particles including massless particles such as photons that compose light. Momentum being a property of particles hints that momentum may have an identity beyond the description of an object's mass multiplied by the object's velocity. Indeed, momentum relates to wave properties and plays a fundamental role in what measurements are taken and how we take these measurements. Furthermore, we find that the conservation of momentum principle is valid when considering systems of particles. We use this principle to analyze the masses and other properties of previously undetected particles, such as the nucleus of an atom and the existence of quarks that make up particles of nuclei. Figure 8.4 below illustrates how a particle scattering backward from another implies that its target is massive and dense. Experiments seeking evidence that quarks make up protons (one type of particle that makes up nuclei) scattered high-energy electrons off of protons (nuclei of hydrogen atoms). Electrons occasionally scattered straight backward in a manner that implied a very small and very dense particle makes up the proton—this observation is considered nearly direct evidence of quarks. The analysis was based partly on the same conservation of momentum principle that works so well on the large scale.

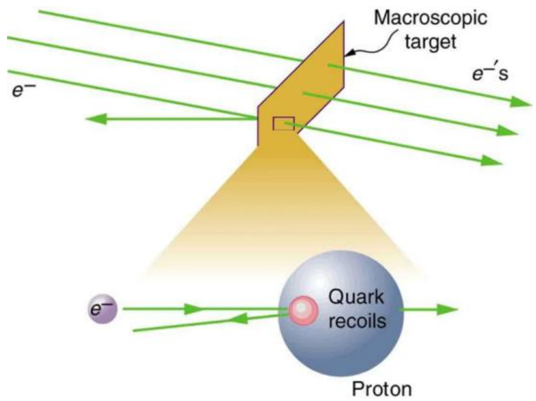


Figure 8.4: A subatomic particle scatters straight backward from a target particle. In experiments seeking evidence for quarks, electrons were observed to occasionally scatter straight backward from a proton.

Let us consider various types of two-object collisions. These collisions are the easiest to analyze, and they illustrate many of the physical principles involved in collisions. The conservation of momentum principle is very useful here, and it can be used whenever the net external force on a system is zero.

We start with the elastic collision of two objects moving along the same line—a one-dimensional problem. An elastic collision is one that also conserves internal kinetic energy. Internal kinetic energy is the sum of the kinetic energies of the objects in the system. Figure 8.5 illustrates an elastic collision in which internal kinetic energy and momentum are conserved.

Truly elastic collisions can only be achieved with subatomic particles, such as electrons striking nuclei. Macroscopic collisions can be very nearly, but not quite, elastic—some kinetic energy is always converted into other forms of energy such as heat transfer due to friction and sound. One macroscopic collision that is nearly elastic is that of two steel blocks on ice. Another nearly elastic collision is that between two carts with spring bumpers on an air track. Icy surfaces and air tracks are nearly frictionless, more readily allowing nearly elastic collisions on them.

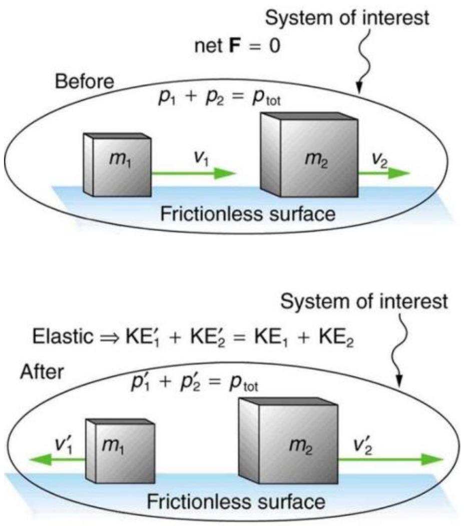


Figure 8.5: An elastic one-dimensional two-object collision. Momentum and internal kinetic energy are conserved.

Now, to solve problems involving one-dimensional elastic collisions between two objects we can use the equations for conservation of momentum and conservation of internal kinetic energy. First, the equation for conservation of momentum for two objects in a one-dimensional collision is,

$$p_1 + p_2 = p'_1 + p'_2 (F_{net} = 0)$$

or

$$m_1 v_1 + m_2 v_2 = m_1 v'_1 + m_2 v'_2 (F_{net} = 0),$$

where the primes (') indicate values after the collision. By definition, an elastic collision conserves internal kinetic energy, and so the sum of kinetic energies before the collision equals the sum after the collision. Thus,

$$½ (m_1 v_1)^2 + ½ (m_2 v_2)^2 = ½ (m_1 v'_1{}^2) + ½ (m_2 v'_2{}^2) \text{ (two-object elastic collision)}$$

expresses the equation for conservation of internal kinetic energy in a one-dimensional collision.

Example - Calculating Velocities Following an Elastic Collision

Calculate the velocities of two objects following an elastic collision, given that

$$m_1 = 0.500 \text{ kg}, m_2 = 3.50 \text{ kg}, v_1 = 4.00 \text{ m/s, and } v_2 = 0.$$

Strategy and Concept

First, visualize what the initial conditions mean—a small object strikes a larger object that is initially at rest. This situation is slightly simpler than the situation shown in Figure 8.5 where both objects are initially moving. We are asked to find two unknowns (the final velocities v'_1 and v'_2). To find two unknowns, we must use two independent equations. Because this collision is elastic, we can use the above two equations. Both can be simplified by the fact that object 2 is initially at rest, and thus $v_2=0$. Once we simplify these equations, we combine them algebraically to solve for the unknowns.

Solution

For this problem, note that $v_2=0$ and use conservation of momentum. Thus,

$$p_1=p'_1+p'_2$$

or

$$m_1v_1 = m_1v'_1+m_2v'_2.$$

Using conservation of internal kinetic energy and that $v_2 = 0$,

$$\tfrac{1}{2} (m_1v_1^2) = \tfrac{1}{2} (m_1v'^2_1) + \tfrac{1}{2} (m_2v'^2_2)$$

Solving the first equation (momentum equation) for v'_2, we obtain

$$v'_2 = [m_1 / m_2] (v_1 - v'_1)$$

Substituting this expression into the second equation (internal kinetic energy equation) eliminates the variable v'_2, leaving only v'_1 as an unknown (the algebra is left as an exercise for the reader). There are two solutions to any quadratic equation; in this example, they are

$$v'_1=4.00 \text{ m/s}$$

and

$$v'_1=-3.00 \text{ m/s}.$$

As noted, when quadratic equations were encountered in earlier chapters, both solutions may or may not be meaningful. In this case, the first solution is the same as the initial condition. The first solution thus represents the situation before the collision and is discarded. The second solution ($v'_1=-3.00$ m/s) is negative, meaning that the first object bounces backward. When this negative value of v'_1 is used to find the velocity of the second object after the collision, we get

$$v'_2 = [m_1 / m_2] (v_1 - v'_1) = [0.500 \text{ kg} / 3.50 \text{ kg}]\ [4.00-(-3.00)] \text{m/s}$$

or

$$v'_2 = 1.00 \text{ m/s}.$$

Discussion

The result of this example is intuitively reasonable. A small object strikes a larger one at rest and bounces backward. The larger one is knocked forward, but with a low speed. (This is like a compact car bouncing backward off a full-size SUV that is initially at rest.) As a check, try calculating the internal kinetic energy before and after the collision. You will see that the internal kinetic energy is unchanged at 4.00 J. Also check the total momentum before and after the collision; you will find it, too, is unchanged.

The equations for conservation of momentum and internal kinetic energy as written above can be used to describe any one-dimensional elastic collision of two objects. These equations can be extended to more objects if needed.

Inelastic Collision

An inelastic collision is one in which the internal kinetic energy changes (it is not conserved).

Figure 8.6 shows an example of an inelastic collision. Two objects that have equal masses head toward one another at equal speeds and then stick together. Their total internal kinetic energy is initially ½ (mv^2)+ ½ (mv^2)=mv^2. The two objects come to rest after sticking together, conserving momentum. But the internal kinetic energy is zero after the collision. A collision in which the objects stick together is sometimes called a perfectly inelastic collision because it reduces internal kinetic energy more than does any other type of inelastic collision. In fact, such a collision reduces internal kinetic energy to the minimum it can have while still conserving momentum.

Perfectly Inelastic Collision

A collision in which the objects stick together is sometimes called "perfectly inelastic."

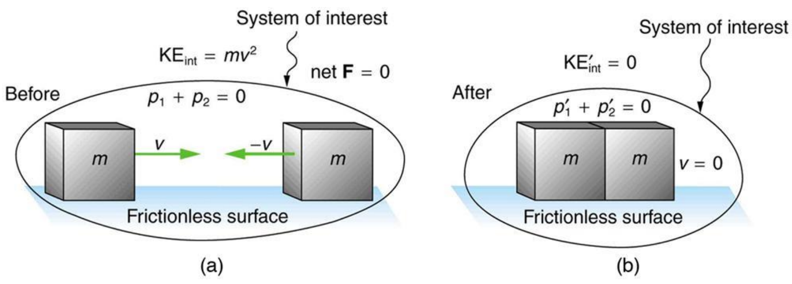


Figure 8.6: An inelastic one-dimensional two-object collision. Momentum is conserved, but internal kinetic energy is not conserved. (a) Two objects of equal mass initially head directly toward one another at the same speed. (b) The objects stick together (a perfectly inelastic

collision), and so their final velocity is zero. The internal kinetic energy of the system changes in any inelastic collision and is reduced to zero in this example.

Example - Calculating Velocity and Change in Kinetic Energy: Inelastic Collision of a Puck and a Goalie

(a) Find the recoil velocity of a 70.0-kg ice hockey goalie, originally at rest, who catches a 0.150-kg hockey puck slapped at him at a velocity of 35.0 m/s. (b) How much kinetic energy is lost during the collision? Assume friction between the ice and the puck-goalie system is negligible.

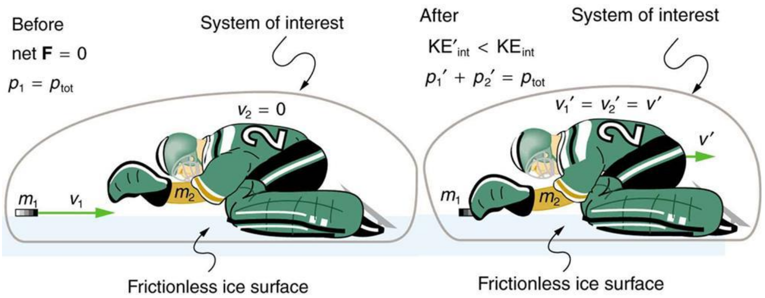


Figure 8.7: An ice hockey goalie catches a hockey puck and recoils backward. The initial kinetic energy of the puck is almost entirely converted to thermal energy and sound in this inelastic collision.

Strategy

Momentum is conserved because the net external force on the puck-goalie system is zero. We can thus use conservation of momentum to find the final velocity of the puck and goalie system. Note that the initial velocity of the goalie is zero and that the final velocity of the puck and goalie are the same. Once the final velocity is found, the kinetic energies can be calculated before and after the collision and compared as requested.

Solution for (a)

Momentum is conserved because the net external force on the puck-goalie system is zero.

Conservation of momentum is

$$p_1+p_2=p'_1+p'_2$$

or

$$m_1v_1+m_2v_2 = m_1v'_1 + m_2v'_2$$

Because the goalie is initially at rest, we know v_2=0. Because the goalie catches the puck, the final velocities are equal, or $v'_1=v'_2=v'$. Thus, the conservation of momentum equation simplifies to

$$m_1v_1=(m_1+m_2)v'.$$

Solving for v' yields

$$v'= [m_1 / (m_1+m_2)]\ v_1.$$

Entering known values in this equation, we get

$$v'= [0.150\text{ kg} / (0.150\text{ kg}+70.0\text{ kg})](35.0\text{ m/s}) = 7.48\times10^{-2}\text{m/s}.$$

Discussion for (a)

This recoil velocity is small and in the same direction as the puck's original velocity, as we might expect.

Solution for (b)

Before the collision, the internal kinetic energy KE_{in}t of the system is that of the hockey puck, because the goalie is initially at rest. Therefore, KE_{int} is initially,

$$\begin{aligned} KE_{int} &= ½ (mv^2) = ½ (0.150\text{ kg})(35.0\text{ m/s})^2 \\ &= 91.9\text{ J}. \end{aligned}$$

After the collision, the internal kinetic energy is

$$\begin{aligned} KE'_{int} &= ½ (m+M)v^2 = ½ (70.15\text{ kg})(7.48\times10^{-2}\text{m/s})^2 \\ &= 0.196\text{ J}. \end{aligned}$$

The change in internal kinetic energy is thus

$$\begin{aligned} KE'_{int} – KE_{int} &= 0.196\text{ J} - 91.9\text{ J} \\ &= -91.7\text{ J} \end{aligned}$$

where the minus sign indicates that the energy was lost.

Discussion for (b)

Nearly all of the initial internal kinetic energy is lost in this perfectly inelastic collision. KE_{int} is mostly converted to thermal energy and sound.

During some collisions, the objects do not stick together and less of the internal kinetic energy is removed—such as happens in most automobile accidents. Alternatively, stored energy may be converted into internal kinetic energy during a collision. Figure 8.8 shows a one-dimensional example in which two carts on an air track collide, releasing potential energy from a compressed spring.

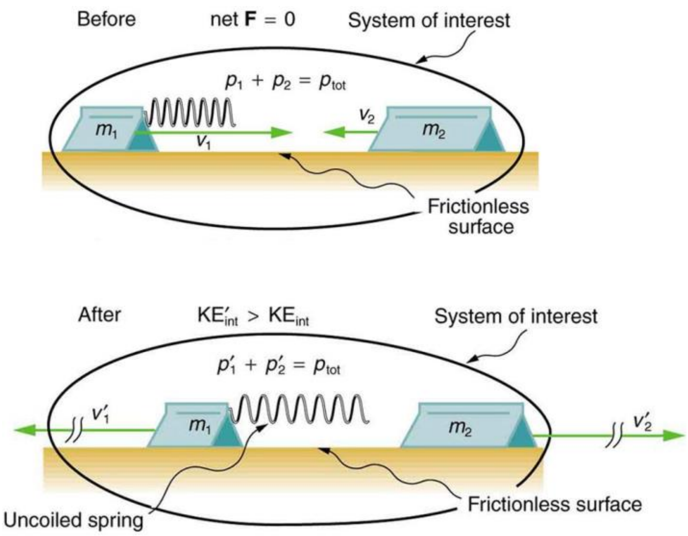


Figure 8.8: An air track is nearly frictionless, so that momentum is conserved. Motion is one-dimensional. In this collision, the potential energy of a compressed spring is released during the collision and is converted to internal kinetic energy.

Collisions are particularly important in sports and the sporting and leisure industry utilizes elastic and inelastic collisions. Let us look briefly at tennis. Recall that in a collision, it is momentum and not force that is important. So, a heavier tennis racquet will have the advantage over a lighter one. This conclusion also holds true for other sports—a lightweight bat (such as a softball bat) cannot hit a hardball very far.

The location of the impact of the tennis ball on the racquet is also important, as is the part of the stroke during which the impact occurs. A smooth motion results in the maximizing of the velocity of the ball after impact and reduces sports injuries such as tennis elbow. A tennis player tries to hit the ball on the “sweet spot” on the racquet, where the vibration and impact are minimized, and the ball is able to be given more velocity. Sports science and technologies also use physics concepts such as momentum and rotational motion and vibrations.

In the previous two sections, we considered only one-dimensional collisions; during such collisions, the incoming and outgoing velocities are all along the same line. But what about collisions, such as those between billiard balls, in which objects scatter to the side? These are two-dimensional collisions, and we shall see that their study is an extension of the one-dimensional analysis already presented. The approach taken (similar to the approach in discussing two-dimensional kinematics and dynamics) is to choose a convenient coordinate system and resolve the motion into components along perpendicular axes. Resolving the motion yields a pair of one-dimensional problems to be solved simultaneously.

One complication arising in two-dimensional collisions is that the objects might rotate before or after their collision. For example, if two ice skaters hook arms as they pass by one another, they will spin in circles. We will not consider such rotation until later, and so for now we arrange things so that no rotation is possible. To avoid rotation, we consider only the scattering of point masses—that is, structureless particles that cannot rotate or spin.

We start by assuming that $\mathbf{F}_{net}=0$, so that momentum **p** is conserved. The simplest collision is one in which one of the particles is initially at rest. (See Figure 8.9.) The best choice for a coordinate system is one with an axis parallel to the velocity of the incoming particle, as shown in Figure 8.9. Because momentum is conserved, the components of momentum along the *x*- and *y*-axes (p_x and p_y) will also be conserved, but with the chosen coordinate system, *py* is initially zero and *px* is the momentum of the incoming particle. Both facts simplify the analysis. (Even with the simplifying assumptions of point masses, one particle initially at rest, and a convenient coordinate system, we still gain new insights into nature from the analysis of two-dimensional collisions.)

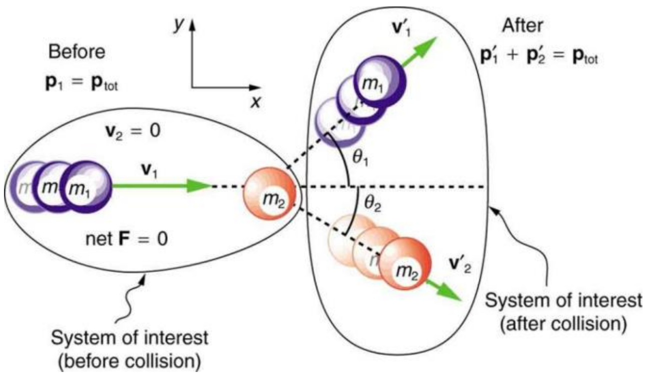


Figure 8.9: A two-dimensional collision with the coordinate system chosen so that *m*2 is initially at rest and *v*1 is parallel to the *x* -axis. This coordinate system is sometimes called the laboratory coordinate system, because many scattering experiments have a target that is stationary in the laboratory, while particles are scattered from it to determine the particles that make-up the target and how they are bound together. The particles may not be observed directly, but their initial and final velocities are.

Along the *x*-axis, the equation for conservation of momentum is,

$$p_{1x} + p_{2x} = p'_{1x} + p'_{2x}$$

Where the subscripts denote the particles and axes and the primes denote the situation after the collision. In terms of masses and velocities, this equation is,

$$m_1v_{1x} + m_2v_{2x} = m_1v'_{1x} + m_2v'_{2x}$$

But because particle 2 is initially at rest, this equation becomes,

$$m_1v_{1x} = m_1v'_{1x} + m_2v'_{2x}$$

The components of the velocities along the x-axis have the form $v \cos\theta$. Because particle 1 initially moves along the x-axis, we find $v_{1x}=v_1$.

Conservation of momentum along the x-axis gives the following equation:

$$m_1v_1=m_1v'_1\cos\theta_1+m_2v'_2\cos\theta_2$$

where θ_1 and θ_2 are as shown in Figure 8.9.

Along the y-axis, the equation for conservation of momentum is

$$p_{1y} + p_{2y} = p'_{1y} + p'_{2y}$$

or

$$m_1v_{1y}+m_2v_{2y} = m_1v'_{1y}+m_2v'_{2y}$$

But v_{1y} is zero, because particle 1 initially moves along the x-axis. Because particle 2 is initially at rest, v_{2y} is also zero. The equation for conservation of momentum along the y-axis becomes,

$$0 = m_1v'_{1y} + m_2v'_{2y}$$

The components of the velocities along the y-axis have the form $v \sin\theta$.

Thus, conservation of momentum along the y-axis gives the following equation:

$$0=m_1v'_1\sin\theta_1+m_2v'_2\sin\theta_2.$$

The equations of conservation of momentum along the x-axis and y-axis are very useful in analyzing two-dimensional collisions of particles, where one is originally stationary (a common laboratory situation). But two equations can only be used to find two unknowns, and so other data may be necessary when collision experiments are used to explore nature at the subatomic level.

8.6 Acceleration of a Rocket

Acceleration of a rocket is

$$a = (ve/m)\ (\Delta m/\Delta t) - g$$

where a is the acceleration of the rocket, ve is the exhaust velocity, m is the mass of the rocket, Δm is the mass of the ejected gas, and Δt is the time in which the gas is ejected.

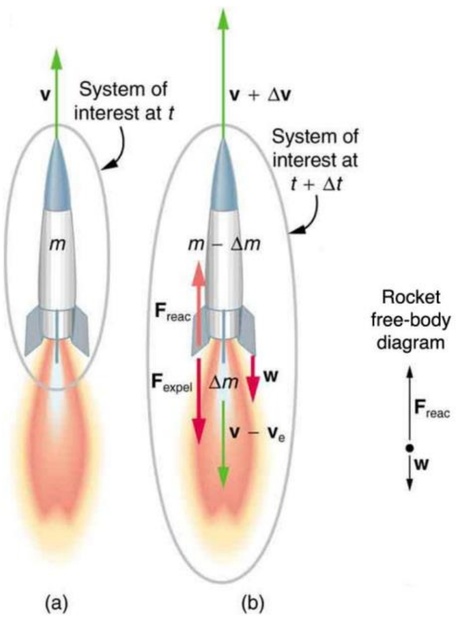


Figure 8.10: (a) This rocket has a mass m and an upward velocity v. The net external force on the system is $-mg$, if air resistance is neglected. (b) A time Δt later the system has two main parts, the ejected gas and the remainder of the rocket. The reaction force on the rocket is what overcomes the gravitational force and accelerates it upward.

A rocket's acceleration depends on three major factors, consistent with the equation for acceleration of a rocket **.** First, the greater the exhaust velocity of the gases relative to the rocket, ve, the greater the acceleration is. The practical limit for ve is about 2.5×10^3m/s for conventional (non-nuclear) hot-gas propulsion systems. The second factor is the rate at which mass is ejected from the rocket. This is the factor $\Delta m/\Delta t$ in the equation. The quantity $(\Delta m/\Delta t)v$e, with units of newtons, is called "thrust." The faster the rocket burns its fuel, the greater its thrust, and the greater its acceleration. The third factor is the mass m of the rocket. The smaller the mass is (all other factors being the same), the greater the acceleration. The rocket mass m decreases dramatically during flight because most of the rocket is fuel to begin with, so that acceleration increases continuously, reaching a maximum just before the fuel is exhausted.

Factors Affecting a Rocket's Acceleration

- The greater the exhaust velocity *ve* of the gases relative to the rocket, the greater the acceleration.
- The faster the rocket burns its fuel, the greater its acceleration.
- The smaller the rocket's mass (all other factors being the same), the greater the acceleration.

For examples and answers, please refer to OpenStax.com questions and answers given on their website or to the College Physics 2e - https://openstax.org/details/books/college-physics-2e

Chapter 9

9.0 Objectives

At the end of this lesson, students should be able to,

- Identify conditions for forces in equilibrium.
- Calculate torque.
- Apply knowledge in rotational force to solve problems.

9.1 Introduction

This chapter covers forces in equilibrium, torque, and relevant calculations.

The first condition necessary to achieve equilibrium is the one already mentioned: the net external force on the system must be zero. Expressed as an equation, this is simply,

$$\text{net } \mathbf{F} = 0$$

Note that if net **F** is zero, then the net external force in *any* direction is zero. For example, the net external forces along the typical *x*- and *y*-axes are zero. This is written as

$$\text{Net } F_x = 0 \text{ and net } F_y = 0$$

Figure 9.1 and Figure 9.2 illustrate situations where net **F**=0 for both static equilibrium (motionless), and dynamic equilibrium (constant velocity).

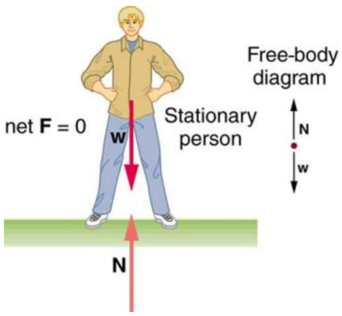


Figure 9.1: This motionless person is in static equilibrium. The forces acting on him add up to zero. Both forces are vertical in this case.

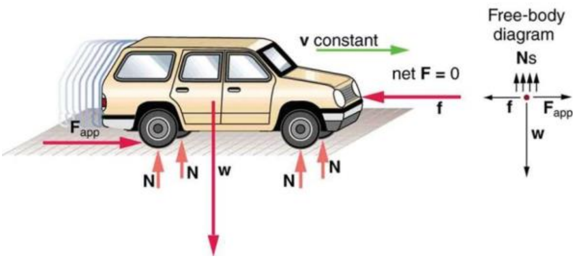


Figure 9.2: This car is in dynamic equilibrium because it is moving at constant velocity. There are horizontal and vertical forces, but the net external force in any direction is zero. The applied force $\mathbf{F}_{app}$ between the tires and the road is balanced by air friction, and the weight of the car is supported by the normal forces, here shown to be equal for all four tires.

However, it is not sufficient for the net external force of a system to be zero for a system to be in equilibrium. Consider the two situations illustrated in Figure 9.3 and Figure 9.4 and where forces are applied to an ice hockey stick lying flat on ice. The net external force is zero in both situations shown in the figure; but in one case, equilibrium is achieved, whereas in the other, it is not. In Figure 9.3, the ice hockey stick remains motionless. But in Figure 9.4, with the same forces applied in different places, the stick experiences accelerated rotation. Therefore, we know that the point at which a force is applied is another factor in determining whether or not equilibrium is achieved. This will be explored further in the next section.

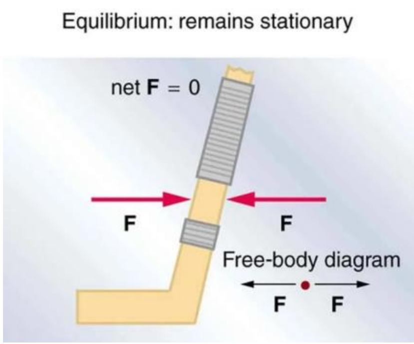


Figure 9.3: An ice hockey stick lying flat on ice with two equal and opposite horizontal forces applied to it. Friction is negligible, and the gravitational force is balanced by the support of the ice (a normal force). Thus, net F = 0. Equilibrium is achieved, which is static equilibrium in this case.

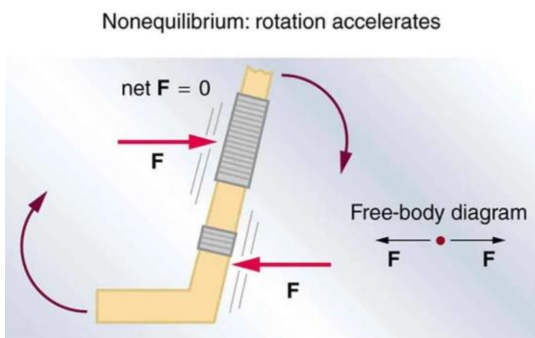


Figure 9.4: The same forces are applied at other points and the stick rotates—in fact, it experiences an accelerated rotation. Here net F = 0 but the system is *not* at equilibrium. Hence, the net F = 0 is a necessary—but not sufficient—condition for achieving equilibrium.

9.2 Torque

The second condition necessary to achieve equilibrium involves avoiding accelerated rotation (maintaining a constant angular velocity). A rotating body or system can be in equilibrium if its rate of rotation is constant and remains unchanged by the forces acting on it. To understand what factors affect rotation, let us think about what happens when you open an ordinary door by rotating it on its hinges.

Several familiar factors determine how effective you are in opening the door. See Figure 9.5. First of all, the larger the force, the more effective it is in opening the door—obviously, the harder you push, the more rapidly the door opens. Also, the point at which you push is crucial. If you apply your force too close to the hinges, the door will open slowly, if at all. Most people have been embarrassed by making this mistake and bumping up against a door when it did not open as quickly as expected. Finally, the direction in which you push is also important. The most effective direction is perpendicular to the door—we push in this direction almost instinctively.

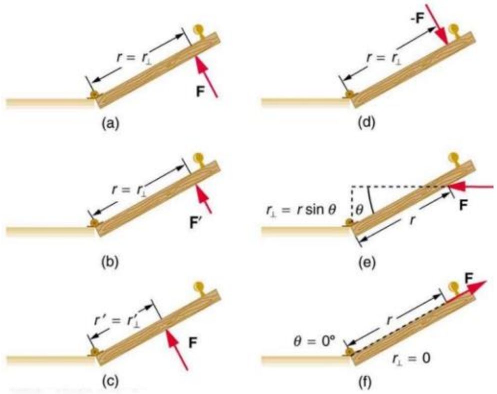


Figure 9.5: Torque is the turning or twisting effectiveness of a force, illustrated here for door rotation on its hinges (as viewed from overhead). Torque has both magnitude and direction. (a) Counterclockwise torque is produced by this force, which means that the door will rotate in a counterclockwise due to **F**. Note that $r\perp$ is the perpendicular distance of the pivot from the line of action of the force. (b) A smaller counterclockwise torque is produced by a smaller force **F′** acting at the same distance from the hinges (the pivot point). (c) The same force as in (a) produces a smaller counterclockwise torque when applied at a smaller distance from the hinges. (d) The same force as in (a), but acting in the opposite direction, produces a clockwise torque. (e) A smaller counterclockwise torque is produced by the same magnitude force acting at the same point but in a different direction. Here, θ is less than 90º. (f) Torque is zero here since the force just pulls on the hinges, producing no rotation. In this case, θ = 0º.

The magnitude, direction, and point of application of the force are incorporated into the definition of the physical quantity called torque. Torque is the rotational equivalent of a force. It is a measure of the effectiveness of a force in changing or accelerating a rotation (changing the angular velocity over a period of time). In equation form, the magnitude of torque is defined to be,

$$\tau = rF\sin\theta$$

where τ (the Greek letter tau) is the symbol for torque, r is the distance from the pivot point to the point where the force is applied, F is the magnitude of the force, and ϑ is the angle between the force and the vector directed from the point of application to the pivot point, as seen in Figure 9.5 and Figure 9.6. An alternative expression for torque is given in terms of the perpendicular lever arm $r\perp$ as shown in Figure 9.5 and Figure 9.6, which is defined as

$$r\perp = r \sin \vartheta$$

so that

$$\tau = r_{\perp}F$$

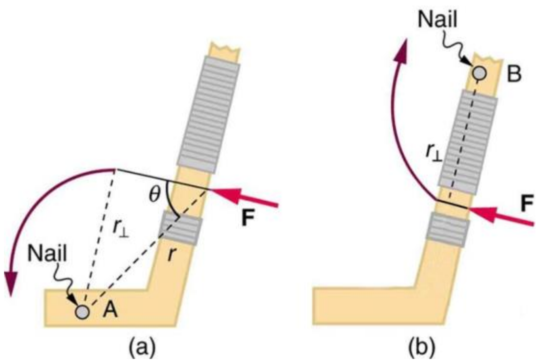


Figure 9.6: A force applied to an object can produce a torque, which depends on the location of the pivot point. (a) The three factors **r**, **F**, and θ for pivot point A on a body are shown here—**r** is the distance from the chosen pivot point to the point where the force **F** is applied, and θ is the angle between **F** and the vector directed from the point of application to the pivot point. If the object can rotate around point A, it will rotate counterclockwise. This means that torque is counterclockwise relative to pivot A. (b) In this case, point B is the pivot point. The torque from the applied force will cause a clockwise rotation around point B, and so it is a clockwise torque relative to B.

The perpendicular lever arm $r\perp$ is the shortest distance from the pivot point to the line along which **F** acts; it is shown as a dashed line in Figure 9.5 and Figure 9.6. Note that the line segment that defines the distance $r\perp$ is perpendicular to **F**, as its name implies. It is sometimes easier to find or visualize $r\perp$ than to find both r and θ. In such cases, it may be more convenient to use $\tau = r\perp\mathbf{F}$ rather than $\tau = r\,F \sin\theta$ for torque, but both are equally valid.

The SI unit of torque is newtons times meters, usually written as N·m. For example, if you push perpendicular to the door with a force of 40 N at a distance of 0.800 m from the hinges, you exert a torque of 32 N·m (0.800 m × 40 N × sin 90º) relative to the hinges. If you reduce the force to 20 N, the torque is reduced to 16 N·m, and so on.

The torque is always calculated with reference to some chosen pivot point. For the same applied force, a different choice for the location of the pivot will give you a different value for the torque, since both r and θ depend on the location of the pivot. Any point in any object can be chosen to calculate the torque about that point. The object may not actually pivot about the chosen "pivot point."

Note that for rotation in a plane, torque has two possible directions. Torque is either clockwise or counterclockwise relative to the chosen pivot point, as illustrated for points B and A, respectively, in Figure 9.6. If the object can rotate about point A, it will rotate counterclockwise, which means that the torque for the force is shown as counterclockwise relative to A. But if the object can rotate about point B, it will rotate clockwise, which means the torque for the force shown is clockwise relative to B. Also, the magnitude of the torque is greater when the lever arm is longer.

Now, *the second condition necessary to achieve equilibrium* is that *the net external torque on a system must be zero*. An external torque is one that is created by an external force. You can choose the point around which the torque is calculated. The point can be the physical pivot point of a system or any other point in space—but it must be the same point for all torques. If the second condition (net external torque on a system is zero) is satisfied for one choice of pivot point, it will also hold true for any other choice of pivot point in or out of the system of interest. (This is true only in an inertial frame of reference.) The second condition necessary to achieve equilibrium is stated in equation form as

$$\text{net } \boldsymbol{\tau} = 0$$

where net means total. Torques, which are in opposite directions are assigned opposite signs. A common convention is to call counterclockwise (ccw) torques positive and clockwise (cw) torques negative.

When two children balance a seesaw as shown in Figure 9.7, they satisfy the two conditions for equilibrium. Most people have perfect intuition about seesaws, knowing that the lighter child must sit farther from the pivot and that a heavier child can keep a lighter one off the ground indefinitely.

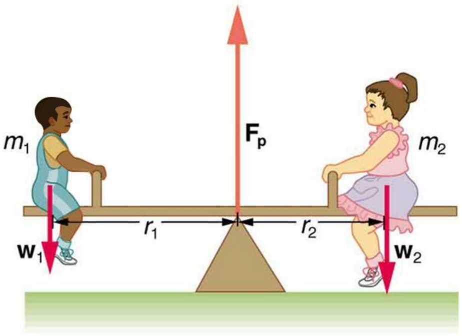


Figure 9.7: Two children balancing a seesaw satisfy both conditions for equilibrium. The lighter child sits farther from the pivot to create a torque equal in magnitude to that of the heavier child.

Example - She Saw Torques on A Seesaw

The two children shown in Figure 9.7 are balanced on a seesaw of negligible mass. (This assumption is made to keep the example simple—more involved examples will follow.) The first child has a mass of 26.0 kg and sits 1.60 m from the pivot. (a) If the second child has a mass of 32.0 kg, how far is she from the pivot? (b) What is F_p, the supporting force exerted by the pivot?

Strategy

Both conditions for equilibrium must be satisfied. In part (a), we are asked for a distance; thus, the second condition (regarding torques) must be used, since the first (regarding only forces) has no distances in it. To apply the second condition for equilibrium, we first identify the system of interest to be the seesaw plus the two children. We take the supporting pivot to be the point about which the torques are calculated. We then identify all external forces acting on the system.

Solution (a)

The three external forces acting on the system are the weights of the two children and the supporting force of the pivot. Let us examine the torque produced by each. Torque is defined to be,

$$\tau = r\,F\,\sin\theta$$

Here $\theta = 90^\circ$, so that $\sin\theta = 1$ for all three forces. That means $r_\perp = r$ for all three. The torques exerted by the three forces are first,

$$\tau_1 = r_1 w_1$$

second,

$$\tau_2 = -r_2 w_2$$

and third,

$$\begin{aligned}\tau_p &= r_p F_p \\ &= 0 \cdot F_p \\ &= 0\end{aligned}$$

Note that a minus sign has been inserted into the second equation because this torque is clockwise and is therefore negative by convention. Since F_p acts directly on the pivot point, the distance r_p is zero. A force acting on the pivot cannot cause a rotation, just as pushing directly on the hinges of a door will not cause it to rotate. Now, the second condition for equilibrium is that the sum of the torques on both children is zero. Therefore

$$\tau_2 = -\tau_1$$

or

$$r_2 w_2 = r_1 w_1$$

Weight is mass times the acceleration due to gravity. Entering mg for w, we get

$$r_2 m_2 g = r_1 m_1 g$$

Solve this for the unknown r_2:

$$r_2 = r_1 (m_1/m_2)$$

The quantities on the right side of the equation are known; thus, r_2 is,

$$r_2 = (1.60 \text{ m}) (26.0 \text{ kg}/32.0 \text{ kg}) = 1.30 \text{ m}$$

As expected, the heavier child must sit closer to the pivot (1.30 m versus 1.60 m) to balance the seesaw.

Solution (b)

This part asks for a force F_p. The easiest way to find it is to use the first condition for equilibrium, which is,

$$\text{net } \mathbf{F} = 0$$

The forces are all vertical, so that we are dealing with a one-dimensional problem along the vertical axis; hence, the condition can be written as

$$\text{net } F_y = 0$$

where we again call the vertical axis the *y*-axis. Choosing upward to be the positive direction, and using plus and minus signs to indicate the directions of the forces, we see that,

$$F_\text{p} - w_1 - w_2 = 0$$

This equation yields what might have been guessed at the beginning:

$$F_\text{p} = w_1 + w_2$$

So, the pivot supplies a supporting force equal to the total weight of the system:

$$F_\text{p} = m_1 g + m_2 g$$

Entering known values gives

$$F\text{p} = (26.0 \text{ kg})(9.80 \text{m/s}^2) + (32.0 \text{ kg})(9.80 \text{m/s}^2)$$

$$= 568 \text{ N}.$$

Discussion

The two results make intuitive sense. The heavier child sits closer to the pivot. The pivot supports the weight of the two children. Part (b) can also be solved using the second condition for equilibrium, since both distances are known, but only if the pivot point is chosen to be somewhere other than the location of the seesaw's actual pivot!

Several aspects of the preceding example have broad implications. First, the choice of the pivot as the point around which torques are calculated simplified the problem. Since F_p is exerted on the pivot point, its lever arm is zero. Hence, the torque exerted by the supporting force F_p is zero relative to that pivot point. The second condition for equilibrium holds for any choice of pivot point, and so we choose the pivot point to simplify the solution of the problem.

Second, the acceleration due to gravity canceled in this problem, and we were left with a ratio of masses. *This will not always be the case*. Always enter the correct forces—do not jump ahead to enter some ratio of masses.

Third, the weight of each child is distributed over an area of the seesaw, yet we treated the weights as if each force were exerted at a single point. This is not an approximation—the distances r_1 and r_2 are the distances to points directly below the center of gravity of each child. As we shall see in the next section, the mass and weight of a system can act as if they are located at a single point.

Finally, note that the concept of torque has an importance beyond static equilibrium. *Torque plays the same role in rotational motion that force plays in linear motion.*

It is one thing to have a system in equilibrium; it is quite another for it to be stable. The toy doll perched on the man's hand in Figure 9.8, for example, is not in stable equilibrium. There are *three types of equilibrium*: *stable*, *unstable*, and *neutral*. Figures throughout this module illustrate various examples.

Figure 9.8 presents a balanced system, such as the toy doll on the man's hand, which has its center of gravity (cg) directly over the pivot, so that the torque of the total weight is zero. This is equivalent to having the torques of the individual parts balanced about the pivot point, in this case the hand. The cgs of the arms, legs, head, and torso are labeled with smaller type.

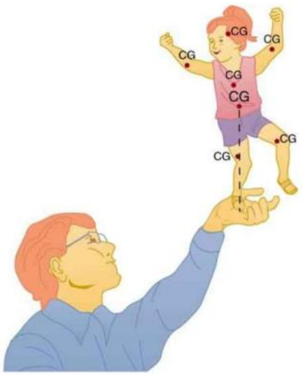


Figure 9.8: A man balances a toy doll on one hand.

A system is said to be in stable equilibrium if, when displaced from equilibrium, it experiences a net force or torque in a direction opposite to the direction of the displacement. For example, a marble at the bottom of a bowl will experience a *restoring* force when displaced from its equilibrium position. This force moves it back toward the equilibrium position. Most systems are in stable equilibrium, especially for small displacements. For another example of stable equilibrium, see the pencil in Figure 9.9.

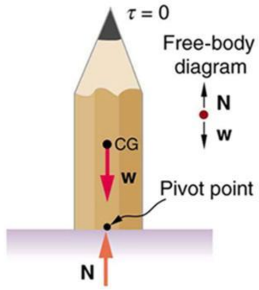


Figure 9.9: This pencil is in the condition of equilibrium. The net force on the pencil is zero and the total torque about any pivot is zero.

A system is in unstable equilibrium if, when displaced, it experiences a net force or torque in the *same* direction as the displacement from equilibrium. A system in unstable equilibrium accelerates away from its equilibrium position if displaced even slightly. An obvious example is a ball resting on top of a hill. Once displaced, it accelerates away from the crest. See the next several figures for examples of unstable equilibrium.

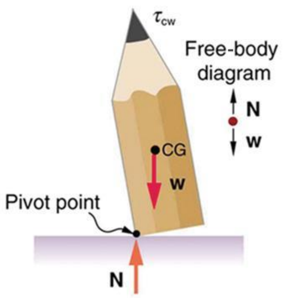


Figure 9.10: If the pencil is displaced slightly to the side (counterclockwise), it is no longer in equilibrium. Its weight produces a clockwise torque that returns the pencil to its equilibrium position.

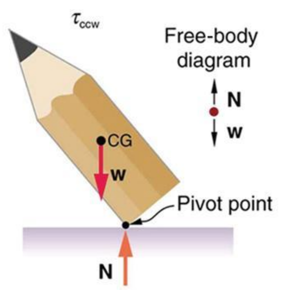


Figure 9.11: If the pencil is displaced too far, the torque caused by its weight changes direction to counterclockwise and causes the displacement to increase.

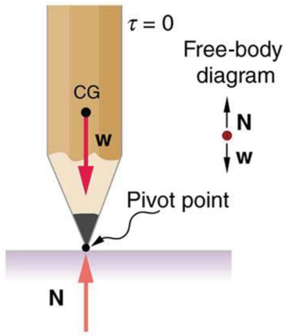


Figure 9.12: This figure shows unstable equilibrium, although both conditions for equilibrium are satisfied.

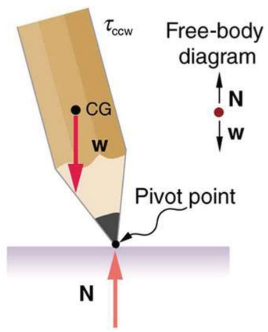


Figure 9.13: If the pencil is displaced even slightly, a torque is created by its weight that is in the same direction as the displacement, causing the displacement to increase.

A system is in neutral equilibrium if its equilibrium is independent of displacements from its original position. A marble on a flat horizontal surface is an example. Combinations of these situations are possible. For example, a marble on a saddle is stable for displacements toward the front or back of the saddle and unstable for displacements to the side. Figure 9.14 shows another example of neutral equilibrium.

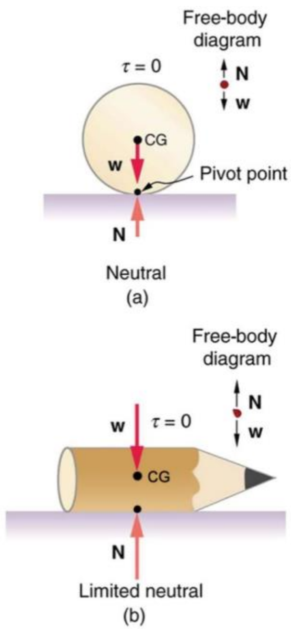


Figure 9.14: (a) Here we see neutral equilibrium. The cg of a sphere on a flat surface lies directly above the point of support, independent of the position on the surface. The sphere is therefore in equilibrium in any location, and if displaced, it will remain put. (b) Because it has a circular cross section, the pencil is in neutral equilibrium for displacements perpendicular to its length.

When we consider how far a system in stable equilibrium can be displaced before it becomes unstable, we find that some systems in stable equilibrium are more stable than others. The pencil in Figure 9.9 and the person in Figure 9.15 (a) are in stable equilibrium, but become unstable for relatively small displacements to the side. The critical point is reached when the cg is no longer *above* the base of support. Additionally, since the cg of a person's body is above the pivots in the hips, displacements must be quickly controlled. This control is a central nervous system function that is developed when we learn to hold our bodies erect as infants. For increased stability while standing, the feet should be spread apart, giving a larger base of support. Stability is also increased by lowering one's center of gravity by bending the knees, as when a football player prepares to receive a ball or braces themselves for a tackle. A cane, a crutch, or a walker increases the stability of the user, even more as the base of support widens. Usually, the cg of a female is lower (closer to the ground) than a male. Young children have their center of gravity between their shoulders, which increases the challenge of learning to walk.

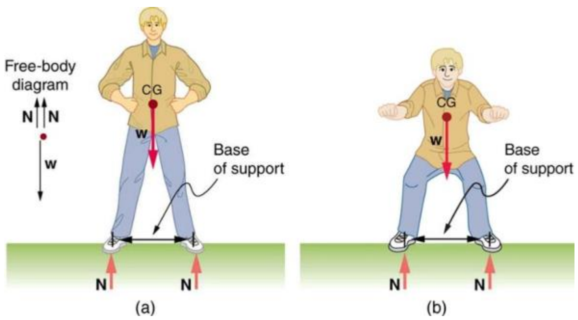


Figure 9.15: (a) The center of gravity of an adult is above the hip joints (one of the main pivots in the body) and lies between two narrowly-separated feet. Like a pencil standing on its eraser, this person is in stable equilibrium in relation to sideways displacements, but relatively small displacements take his cg outside the base of support and make him unstable. Humans are less stable relative to forward and backward displacements because the feet are not very long. Muscles are used extensively to balance the body in the front-to-back direction. (b) While bending in the manner shown, stability is increased by lowering the center of gravity. Stability is also increased if the base is expanded by placing the feet farther apart.

Animals such as chickens have easier systems to control. Figure 9.16 shows that the cg of a chicken lies below its hip joints and between its widely separated and broad feet. Even relatively large displacements of the chicken's cg are stable and result in restoring forces and torques that return the cg to its equilibrium position with little effort on the chicken's part. Not all birds are like chickens, of course. Some birds, such as the flamingo, have balance systems that are almost as sophisticated as that of humans.

Figure 9.16: shows that the cg of a chicken is below the hip joints and lies above a broad base of support formed by widely-separated and large feet. Hence, the chicken is in very stable equilibrium, since a relatively large displacement is needed to render it unstable. The body of the chicken is supported from above by the hips and acts as a pendulum between the hips. Therefore, the chicken is stable for front-to-back displacements as well as for side-to-side displacements.

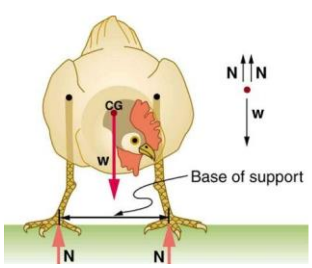


Figure 9.16: The center of gravity of a chicken is below the hip joints. The chicken is in stable equilibrium. The body of the chicken is supported from above by the hips and acts as a pendulum between them.

Engineers and architects strive to achieve extremely stable equilibriums for buildings and other systems that must withstand wind, earthquakes, and other forces that displace them from equilibrium. Although the examples in this section emphasize gravitational forces, the basic conditions for equilibrium are the same for all types of forces. The net external force must be zero, and the net torque must also be zero. Statics can be applied to a variety of situations, ranging from raising a drawbridge to bad posture and back strain.

For examples and answers, please refer to OpenStax.com questions and answers given on their website or to the College Physics 2e - https://openstax.org/details/books/college-physics-2e

Chapter 11

11.0 Objectives

At the end of this lesson, students should be able to,

- Identify different types of fluids.
- Identify cohesion and adhesion.
- Apply Pascal's principle.
- Apply Archimedes' principle.
- Calculate density and pressure of fluids.

11.1 Introduction

This chapter covers fluids, cohesion and adhesion bonds, Pascal’s and Archimedes’ principles and related calculations.

Matter most commonly exists as a solid, liquid, gas, or plasma; these states are known as the common *phases of matter*. Solids have a definite shape and a specific volume, liquids have a definite volume but their shape changes depending on the container in which they are held, gases have neither a definite shape nor a specific volume as their molecules move to fill the container in which they are held, and plasmas also have neither definite shape nor volume. (See Figure 11.1.) Liquids, gases, and plasmas are considered to be fluids because they yield to shearing forces, whereas solids resist them. Note that the extent to which fluids yield to shearing forces (and hence flow easily and quickly) depends on a quantity called the viscosity. We can understand the phases of matter and what constitutes a fluid by considering the forces between atoms that make up matter in the three phases.

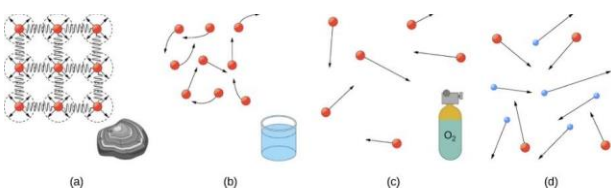


Figure 11.1: (a) Atoms in a solid always have the same neighbors, held near home by forces represented here by springs. These atoms are essentially in contact with one another. A rock is an example of a solid. This rock retains its shape because of the forces holding its atoms together. (b) Atoms in a liquid are also in close contact but can slide over one another. Forces between them strongly resist attempts to push them closer together and also hold them in close contact. Water is an example of a liquid. Water can flow, but it also remains in an open container because of the forces between its atoms. (c) Atoms in a gas are separated by distances that are considerably larger than the size of the atoms themselves, and they move about freely. A gas must be held in a closed container to prevent it from moving out freely. (d) A plasma is composed of electrons, protons, and ions that, like gases, are spaced far apart and move about freely.

Atoms in *solids* are in close contact, with forces between them that allow the atoms to vibrate but not to change positions with neighboring atoms. (These forces can be thought of as springs that can be stretched or compressed, but not easily broken.) Thus a solid *resists* all types of stress. A solid cannot be easily deformed because the atoms that make up the solid are not able to move about freely. Solids also resist compression, because their atoms form part of a lattice structure in which the atoms are a relatively fixed distance apart. Under compression, the atoms would be forced into one another. Most of the examples we have studied so far have involved solid objects which deform very little when stressed.

In contrast, *liquids* deform easily when stressed and do not spring back to their original shape once the force is removed because the atoms are free to slide about and change neighbors—that is, they *flow* (so they are a type of fluid), with the molecules held together by their mutual attraction. When a liquid is placed in a container with no lid on, it remains in the container (providing the container has no holes below the surface of the liquid!). Because the atoms are closely packed, liquids, like solids, resist compression.

Atoms in *gases* and charged particles in *plasmas* are separated by distances that are large compared with the size of the particles. The forces between the particles are therefore very weak, except when they collide with one another. Gases and plasmas thus not only flow (and are therefore considered to be fluids) but they are relatively easy to compress because there is much space and little force between the particles. When placed in an open container, gases, unlike liquids, will escape. The major distinction is that gases are easily compressed, whereas liquids are not. Plasmas are difficult to contain because they have so much energy. When discussing how substances flow, we shall generally refer to both gases and liquids simply as fluids and make a distinction between them only when they behave differently.

Density is an important characteristic of substances. It is crucial, for example, in determining whether an object sinks or floats in a fluid. Density is the mass per unit volume of a substance or object. In equation form, density is defined as

$$\rho = m/V$$

where the Greek letter ρ (rho) is the symbol for density, m is the mass, and V is the volume occupied by the substance.

11.2 Density

Density is mass per unit volume.

$$\rho = m/V$$

where ρ is the symbol for density, m is the mass, and V is the volume occupied by the substance.

In the riddle regarding the feathers and bricks, the masses are the same, but the volume occupied by the feathers is much greater, since their density is much lower. The SI unit of density is kg/m^3, representative values are given in Table 11.1. The metric system was originally devised so that water would have a density of 1g/cm3, equivalent to 10^3kg/m^3. Thus, the basic mass unit, the kilogram, was first devised to be the mass of 1000 mL of water, which has a volume of 1000 cm^3.

Substance	**ρ(×10^3kg/m^3 or g/mL)**	**Substance**	**ρ(10^3kg/m^3 or g/mL)**	**Substance**	**ρ(10^3kg/m^3 or g/mL)**
Solids		**Liquids**		**Gases**	
Aluminum	2.7	Water (4ºC)	1.000	Air	1.29×10−3
Brass	8.44	Blood	1.05	Carbon dioxide	1.98×10−3
Copper (average)	8.8	Sea water	1.025	Carbon monoxide	1.25×10−3
Gold	19.32	Mercury	13.6	Hydrogen	0.090×10−3
Iron or steel	7.8	Ethyl alcohol	0.79	Helium	0.18×10−3
Lead	11.3	Gasoline	0.68	Methane	0.72×10−3
Polystyrene	0.10	Glycerin	1.26	Nitrogen	1.25×10−3
Tungsten	19.30	Olive oil	0.92	Nitrous oxide	1.98×10−3
Uranium	18.70			Oxygen	1.43×10−3
Concrete	2.30–3.0			Steam (100º C)	0.60×10−3
Cork	0.24				
Glass, common (average)	2.6				
Granite	2.7				
Earth's crust	3.3				
Wood	0.3–0.9				
Ice (0°C)	0.917				
Bone	1.7–2.0				
Silver	10.49				

Table 11.1: Densities of Various Substances

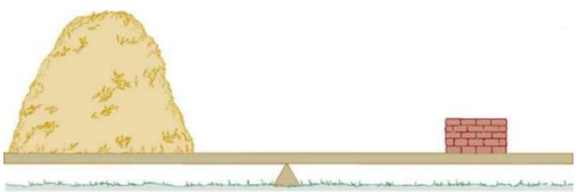


Figure 11.2: A ton of feathers and a ton of bricks have the same mass, but the feathers make a much bigger pile because they have a much lower density.

As you can see by examining Table 11.1, the density of an object may help identify its composition. The density of gold, for example, is about 2.5 times the density of iron, which is about 2.5 times the density of aluminum. Density also reveals something about the phase of the matter and its substructure. Notice that the densities of liquids and solids are roughly comparable, consistent with the fact that their atoms are in close contact. The densities of gases are much less than those of liquids and solids, because the atoms in gases are separated by large amounts of empty space.

Example - Calculating the Mass of a Reservoir From Its Volume

A reservoir has a surface area of 50.0km^2 and an average depth of 40.0 m. What mass of water is held behind the dam? (See Figure 11.3 for a view of a large reservoir—the Three Gorges Dam site on the Yangtze River in central China.)

Strategy

We can calculate the volume V of the reservoir from its dimensions and find the density of water ρ in Table 11.1. Then the mass m can be found from the definition of density

$$\rho = m/V$$

Solution

Solving equation $\rho = m/V$ for m gives $m = \rho V$

The volume V of the reservoir is its surface area A times its average depth h:

$$V = Ah = (50.0\text{km}^2)(40.0\text{m})$$

$$= [(50.0\text{ km}^2)(10^3\text{m}/1\text{km})^2](40.0\text{ m})$$

$$= 2.00\times10^9\text{m}^3$$

The density of water ρ from Table 11.1 is $1.000\times10^3\text{kg/m}^3$. Substituting V and ρ into the expression for mass gives

$$M = (1.00\times10^3\text{kg/m}^3)(2.00\times10^9\text{m}^3)$$
$$= 2.00\times10^{12}\text{kg}.$$

Discussion

A large reservoir contains a very large mass of water. In this example, the weight of the water in the reservoir is $mg = 1.96\times10^{13}\text{N}$, where g is the acceleration due to the Earth's gravity (about 9.80m/s^2). It is reasonable to ask whether the dam must supply a force equal to this tremendous weight. The answer is no. As we shall see in the following sections, the force the dam must supply can be much smaller than the weight of the water it holds back.

Figure 11.3: Three Gorges Dam in central China. When completed in 2008, this became the world's largest hydroelectric plant, generating power equivalent to that generated by 22 average-sized nuclear power plants. The concrete dam is 181 m high and 2.3 km across. The reservoir made by this dam is 660 km long. Over 1 million people were displaced by the creation of the reservoir. (credit: Le Grand Portage)

11.3 Pressure

Pressure is defined as the force divided by the area perpendicular to the force over which the force is applied, or

$$P = F/A$$

A given force can have a significantly different effect depending on the area over which the force is exerted, as shown in Figure 11.4. The SI unit for pressure is the *pascal*, where

$$1\text{ Pa}=1\text{N/m}^2$$

In addition to the pascal, there are many other units for pressure that are in common use. In meteorology, atmospheric pressure is often described in units of millibar (mb), where

$$100\text{ mb}=1\times10^4\text{Pa}$$

Pounds per square inch (lb/in^2 or psi) is still sometimes used as a measure of tire pressure, and millimeters of mercury (mm Hg) is still often used in the measurement of blood pressure. Pressure is defined for all states of matter but is particularly important when discussing fluids.

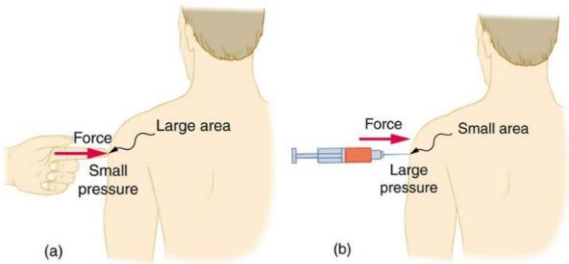


Figure 11.4: (a) While the person being poked with the finger might be irritated, the force has little lasting effect. (b) In contrast, the same force applied to an area the size of the sharp end of a needle is great enough to break the skin.

Example - Calculating Force Exerted by the Air: What Force Does a Pressure Exert?

An astronaut is working outside the International Space Station where the atmospheric pressure is essentially zero. The pressure gauge on her air tank reads 6.90×10^6Pa. What force does the air inside the tank exert on the flat end of the cylindrical tank, a disk 0.150 m in diameter?

Strategy

We can find the force exerted from the definition of pressure given in $P = F/A$, provided we can find the area A acted upon.

Solution

By rearranging the definition of pressure to solve for force, we see that,

$$F = PA.$$

Here, the pressure P is given, as is the area of the end of the cylinder A, given by $A=\pi r2$. Thus,

$$F = (6.90\times10^6\text{N/m}^2)(3.14)(0.0750\text{ m})^2$$

$$= 1.22\times10^5\text{N}.$$

Discussion

Wow! No wonder the tank must be strong. Since we found $F = PA$, we see that the force exerted by a pressure is directly proportional to the area acted upon as well as the pressure itself.

The force exerted on the end of the tank is perpendicular to its inside surface. This direction is because the force is exerted by a static or stationary fluid. We have already seen that fluids cannot *withstand* shearing (sideways) forces; they cannot *exert* shearing forces, either. Fluid pressure has no direction, being a scalar quantity. The forces due to pressure have well-defined directions: they are always exerted perpendicular to any surface. (See the tire in Figure 11. 5 for example.) Finally, note that pressure is exerted on all surfaces. Swimmers, as well as the tire, feel pressure on all sides. (See Figure 11.6)

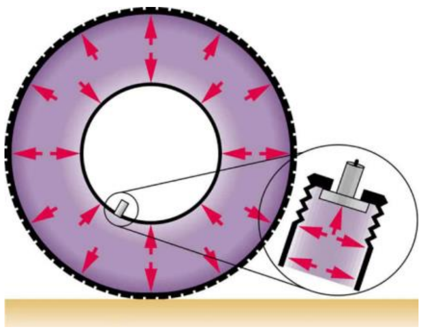


Figure 11.5: Pressure inside this tire exerts forces perpendicular to all surfaces it contacts. The arrows give representative directions and magnitudes of the forces exerted at various points. Note that static fluids do not exert shearing forces.

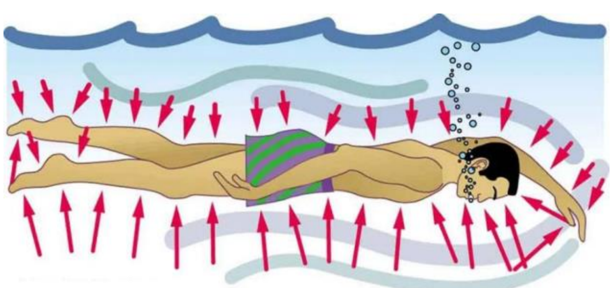


Figure 11.6: Pressure is exerted on all sides of this swimmer, since the water would flow into the space he occupies if he were not there. The arrows represent the directions and magnitudes of the forces exerted at various points on the swimmer. Note that the forces are larger underneath, due to greater depth, giving a net upward or buoyant force that is balanced by the weight of the swimmer.

If your ears have ever popped on a plane flight or ached during a deep dive in a swimming pool, you have experienced the effect of depth on pressure in a fluid. At the Earth's surface, the air pressure exerted on you is a result of the weight of air above you. This pressure is reduced as you climb up in altitude and the weight of air above you decreases. Underwater, the pressure exerted on you increases with increasing depth. In this case, the pressure being exerted upon you is a result of both the weight of water above you *and* that of the atmosphere above you. You may notice an air pressure change on an elevator ride that transports you many stories, but you need only dive a meter or so below the surface of a pool to feel a pressure increase. The difference is that water is much denser than air, about 775 times as dense.

Consider the container in Figure 11.7. Its bottom supports the weight of the fluid in it. Let us calculate the pressure exerted on the bottom by the weight of the fluid. That pressure is the weight of the fluid mg divided by the area A supporting it (the area of the bottom of the container):

$$P = mg/A$$

We can find the mass of the fluid from its volume and density:

$$m = \rho V$$

The volume of the fluid V is related to the dimensions of the container. It is,

$$V = Ah$$

where A is the cross-sectional area and h is the depth. Combining the last two equations gives

$$m = \rho Ah$$

If we enter this into the expression for pressure, we obtain,

$$P = (\rho Ah)g/ A$$

The area cancels, and rearranging the variables yields,

$$P = h\rho g$$

This value is the *pressure due to the weight of a fluid*. The equation has general validity beyond the special conditions under which it is derived here. Even if the container were not there, the surrounding fluid would still exert this pressure, keeping the fluid static. Thus, the equation $P = h\rho g$ represents the pressure due to the weight of any fluid of *average density* ρ at any depth h below its surface. For liquids, which are nearly incompressible, this equation holds to great depths. For gases, which are quite compressible, one can apply this equation as long as the density changes are small over the depth considered.

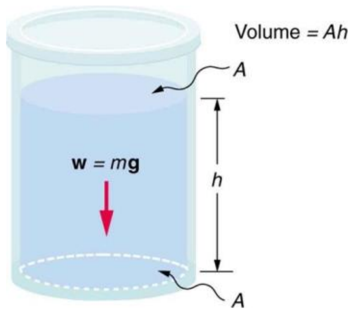


Figure 11.7: The bottom of this container supports the entire weight of the fluid in it. The vertical sides cannot exert an upward force on the fluid (since it cannot withstand a shearing force), and so the bottom must support it all.

Example - Calculating the Average Pressure and Force Exerted: What Force Must a Dam Withstand?

In a previous example, we calculated the mass of water in a large reservoir. We will now consider the pressure and force acting on the dam retaining water. (See Figure 11.8.) The dam is 500 m wide, and the water is 80.0 m deep at the dam. (a) What is the average pressure on the dam due to the water? (b) Calculate the force exerted against the dam and compare it with the weight of water in the dam (previously found to be 1.96×10^{13}N).

Strategy for (a)

The average pressure P_{av} due to the weight of the water is the pressure at the average depth h_{av} of 40.0 m, since pressure increases linearly with depth.

Solution for (a)

The average pressure due to the weight of a fluid is,

$$P_{\text{av}} = h_{\text{av}}\rho \text{g}$$

Entering the density of water from Table 11.1 and taking h_{av} to be the average depth of 40.0 m, we obtain

$$P_{av} = (40.0\text{ m})(10^3\text{kg/m}^3)(9.80\text{m/s}^2)$$

$$= 3.92\times10^5\text{ N/m}^2$$

$$= 392\text{ kPa.}$$

Strategy for (b)

The force exerted on the dam by the water is the average pressure times the area of contact:

$$F = P_{av}A$$

Solution for (b)

We have already found the value for P_{av}. The area of the dam is A = 80.0 m × 500 m = 4.00 × 10^4 m^2, so that

$$F = (3.92\times10^5 \text{N/m}^2)(4.00\times10^4 \text{m}^2)$$

$$= 1.57\times10^{10} \text{N}$$

Discussion

Although this force seems large, it is small compared with the 1.96×10^{13}N weight of the water in the reservoir—in fact, it is only 0.0800% of the weight. Note that the pressure found in part (a) is completely independent of the width and length of the lake—it depends only on its average depth at the dam. Thus, the force depends only on the water's average depth and the dimensions of the dam, *not* on the horizontal extent of the reservoir. In the diagram, the thickness of the dam increases with depth to balance the increasing force due to the increasing pressure.

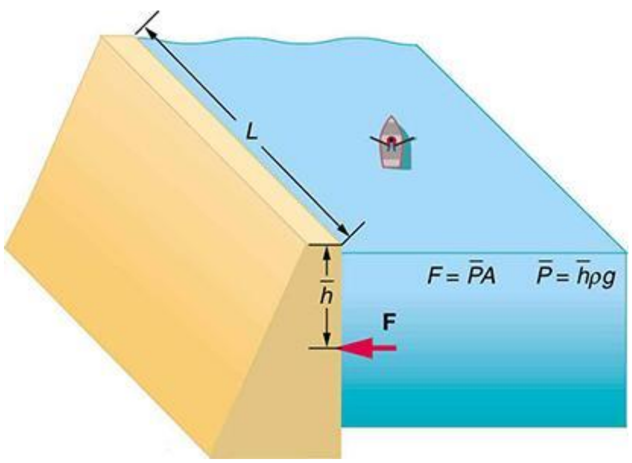


Figure 11.8: The dam must withstand the force exerted against it by the water it retains. This force is small compared with the weight of the water behind the dam.

Atmospheric pressure is another example of pressure due to the weight of a fluid, in this case due to the weight of *air* above a given height. The atmospheric pressure at the Earth's surface varies a little due to the large-scale flow of the atmosphere induced by the Earth's rotation (this creates weather "highs" and "lows"). However, the average pressure at sea level is given by the *standard atmospheric pressure* P_{atm}, measured to be,

$$1 \text{ atmosphere (atm)} = P_{atm} = 1.01 \times 10^5 \text{N/m}^2 = 101 \text{ kPa}.$$

This relationship means that, on average, at sea level, a column of air above 1.00m^2 of the Earth's surface has a weight of $1.01 \times 10^5 \text{N}$, equivalent to 1 atm. (See Figure 11.9)

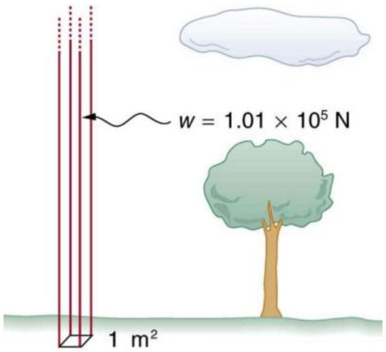


Figure 11.9: Atmospheric pressure at sea level averages 1.01×105Pa (equivalent to 1 atm), since the column of air over this 1m^2, extending to the top of the atmosphere, weighs 1.01×105N.

11.4 Pascal's Principle

A change in pressure applied to an enclosed fluid is transmitted undiminished to all portions of the fluid and to the walls of its container.

Pascal's principle, an experimentally verified fact, is what makes pressure so important in fluids. Since a change in pressure is transmitted undiminished in an enclosed fluid, we often know more about pressure than other physical quantities in fluids. Moreover, Pascal's principle implies that *the total pressure in a fluid is the sum of the pressures from different sources*. We shall find this fact—that pressures add—very useful.

Application of Pascal's Principle

One of the most important technological applications of Pascal's principle is found in a *hydraulic system*, which is an enclosed fluid system used to exert forces. The most common hydraulic systems are those that operate car brakes. Let us first consider the simple hydraulic system shown in Figure 11.10.

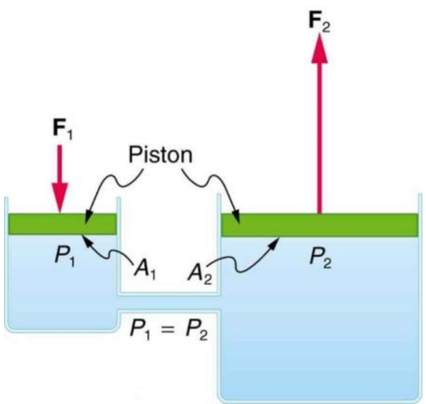


Figure 11.10: A typical hydraulic system with two fluid-filled cylinders, capped with pistons and connected by a tube called a hydraulic line. A downward force $\mathbf{F}_1$ on the left piston creates a pressure that is transmitted undiminished to all parts of the enclosed fluid. This results in an upward force $\mathbf{F}_2$ on the right piston that is larger than $\mathbf{F}_1$ because the right piston has a larger area.

Relationship Between Forces in a Hydraulic System

We can derive a relationship between the forces in the simple hydraulic system shown in Figure 11.10 by applying Pascal's principle. Note first that the two pistons in the system are at the same height, and so there will be no difference in pressure due to a difference in depth. Now the pressure due to F_1 acting on area A_1 is simply $P_1=F_1/A_1$, as defined by $P=/FA$. According to Pascal's principle, this pressure is transmitted undiminished throughout the fluid and to all walls of the container. Thus, a pressure P_2 is felt at the other piston that is equal to P_1. That is $P_1=P_2$.

But since $P_2 = F_2/A_2$, we see that $F_1/A_1=F_2/A_2$.

This equation relates the ratios of force to area in any hydraulic system, providing the pistons are at the same vertical height and that friction in the system is negligible. Hydraulic systems can increase or decrease the force applied to them. To make the force larger, the pressure is applied to a larger area. For example, if a 100-N force is applied to the left cylinder in Figure 11.10 and the right one has an area five times greater, then the force out is 500 N. Hydraulic systems are analogous to simple levers, but they have the advantage that pressure can be sent through tortuously curved lines to several places at once.

Example - Calculating Force of Wheel Cylinders: Pascal Puts on the Brakes

Consider the automobile hydraulic system shown in Figure 11.11.

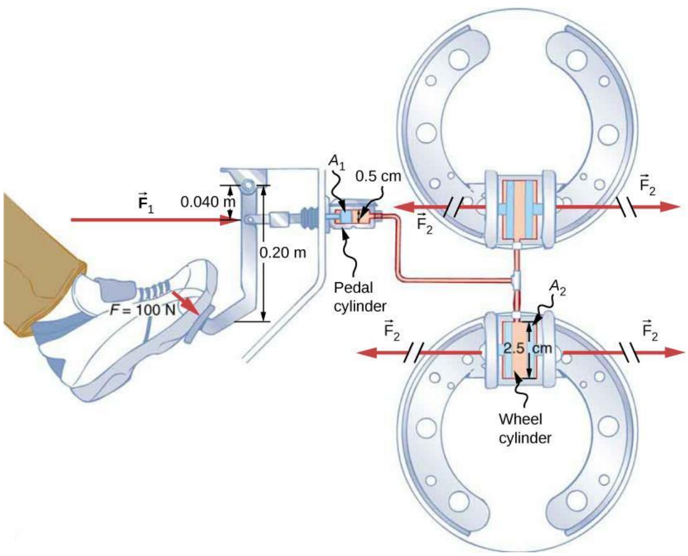


Figure 11.11: Hydraulic brakes use Pascal's principle. The driver exerts a force of 100 N on the brake pedal. This force is increased by the simple lever and again by the hydraulic system. Each of the identical wheel cylinders receives the same pressure and, therefore, creates the same force output F_2. The circular cross-sectional areas of the pedal and wheel cylinders are represented by A_1 and A_2, respectively.

A force of 100 N is applied to the brake pedal, which acts on the pedal cylinder through a lever. A force of 500 N is exerted on the pedal cylinder. Pressure created in the pedal cylinder is transmitted to four-wheel cylinders. The pedal cylinder has a diameter of 0.500 cm, and each wheel cylinder has a diameter of 2.50 cm. Calculate the force F_2 created at each of the wheel cylinders.

Strategy

We are given the force $F1$ that is applied to the pedal cylinder. The cross-sectional areas A_1 and A_2 can be calculated from their given diameters. Then $F_1A_1 = F_2A_2$ can be used to find the force F_2. Manipulate this algebraically to get F_2 on one side and substitute known values:

Solution

Pascal's principle applied to hydraulic systems is given by $F_1A_1 = F_2A_2$:

$$F_2 = (A_2/A_1)F1 = (\pi r_2^2 / \pi r^2_1\ )F1 = [(1.25\text{ cm})^2 / (0.250\text{ cm})^2]\ 500\text{ N}$$

$$= 1.25\times10^4\text{N}.$$

Discussion

This value is the force exerted by each of the four-wheel cylinders. Note that we can add as many wheel cylinders as we wish. If each has a 2.50-cm diameter, each will exert 1.25×10^4N.

A simple hydraulic system, such as a simple machine, can increase force but cannot do more work than done on it. Work is force times distance moved, and the wheel cylinder moves through a smaller distance than the pedal cylinder. Furthermore, the more wheels added, the smaller the distance each moves. Many hydraulic systems—such as power brakes and those in bulldozers—have a motorized pump that actually does most of the work in the system. The movement of the legs of a spider is achieved partly by hydraulics. Using hydraulics, a jumping spider can create a force that makes it capable of jumping 25 times its length!

If you limp into a gas station with a nearly flat tire, you will notice the tire gauge on the airline reads nearly zero when you begin to fill it. In fact, if there were a gaping hole in your tire, the gauge would read zero, even though atmospheric pressure exists in the tire. Why does the gauge read zero? There is no mystery here. Tire gauges are simply designed to read zero at atmospheric pressure and positive when pressure is greater than atmospheric.

Similarly, atmospheric pressure adds to blood pressure in every part of the circulatory system. (As noted in Pascal's Principle, the total pressure in a fluid is the sum of the pressures from different sources—here, the heart and the atmosphere.) But atmospheric pressure has no net effect on blood flow since it adds to the pressure coming out of the heart and going back into it, too. What is important is how much *greater* blood pressure is than atmospheric pressure. Blood pressure measurements, like tire pressures, are thus made relative to atmospheric pressure.

In brief, it is very common for pressure gauges to ignore atmospheric pressure—that is, to read zero at atmospheric pressure. We therefore define gauge pressure to be the pressure relative to atmospheric pressure. Gauge pressure is positive for pressures above atmospheric pressure, and negative for pressures below it.

11.5 Gauge Pressure

Gauge pressure is the pressure relative to atmospheric pressure. Gauge pressure is positive for pressures above atmospheric pressure, and negative for pressures below it.

In fact, atmospheric pressure does add to the pressure in any fluid not enclosed in a rigid container. This happens because of Pascal's principle. The total pressure, or absolute pressure, is thus the sum of gauge pressure and atmospheric pressure: $P_{\text{abs}} = P_{\text{g}} + P_{\text{atm}}$ where P_{abs} is absolute pressure, P_{g} is gauge pressure, and P_{atm} is atmospheric pressure. For example, if your tire gauge reads 34 psi (pounds per square inch), then the absolute pressure is 34 psi plus 14.7 psi (P_{atm} in psi), or 48.7 psi (equivalent to 336 kPa).

Absolute Pressure

Absolute pressure is the sum of gauge pressure and atmospheric pressure.

For reasons we will explore later, in most cases the absolute pressure in fluids cannot be negative. Fluids push rather than pull, so the smallest absolute pressure is zero. (A negative absolute pressure is a pull.) Thus, the smallest possible gauge pressure is $P_g = -P_{atm}$ (this makes P_{abs} zero). There is no theoretical limit to how large a gauge pressure can be.

There are a host of devices for measuring pressure, ranging from tire gauges to blood pressure cuffs. Pascal's principle is of major importance in these devices. The undiminished transmission of pressure through a fluid allows precise remote sensing of pressures. Remote sensing is often more convenient than putting a measuring device into a system, such as a person's artery.

Figure 11.12 shows one of the many types of mechanical pressure gauges in use today. In all mechanical pressure gauges, pressure results in a force that is converted (or transduced) into some type of readout.

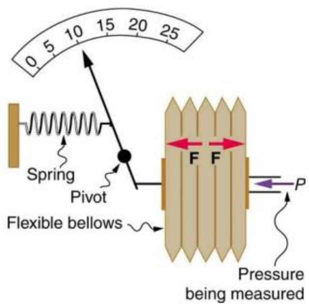


Figure 11.12: This aneroid gauge utilizes flexible bellows connected to a mechanical indicator to measure pressure.

An entire class of gauges uses the property that pressure due to the weight of a fluid is given by $P = h\rho g$. Consider the U-shaped tube shown in Figure 11.13, for example. This simple tube is called a *manometer*. In Figure 11.13 (a), both sides of the tube are open to the atmosphere. Atmospheric pressure therefore pushes down on each side equally so its effect cancels. If the fluid is deeper on one side, there is a greater pressure on the deeper side, and the fluid flows away from that side until the depths are equal.

Let us examine how a manometer is used to measure pressure. Suppose one side of the U-tube is connected to some source of pressure P_{abs} such as the toy balloon in Figure 11.13 (b) or the vacuum-packed peanut jar shown in Figure 11.3 (c). Pressure is transmitted undiminished to the manometer, and the fluid levels are no longer equal. In Figure 11.3(b), P_{abs} is greater than atmospheric pressure, whereas in Figure 11.13 (c), P_{abs} is less than atmospheric pressure. In both cases, P_{abs} differs from atmospheric pressure by an amount $h\rho g$, where ρ is the density of the fluid in the manometer. In Figure 11.13 (b), P_{abs} can support a column of fluid of height h, and so it must exert a pressure $h\rho g$ greater than atmospheric pressure (the gauge pressure P_g is positive). In Figure 11.13 (c), atmospheric pressure can support a column of fluid of height h, and so P_{abs} is less than atmospheric pressure by an amount $h\rho g$ (the gauge pressure P_g is negative). A manometer with one side open to the atmosphere is an ideal device for measuring gauge pressures. The gauge pressure is $P_g = h\rho g$ and is found by measuring h.

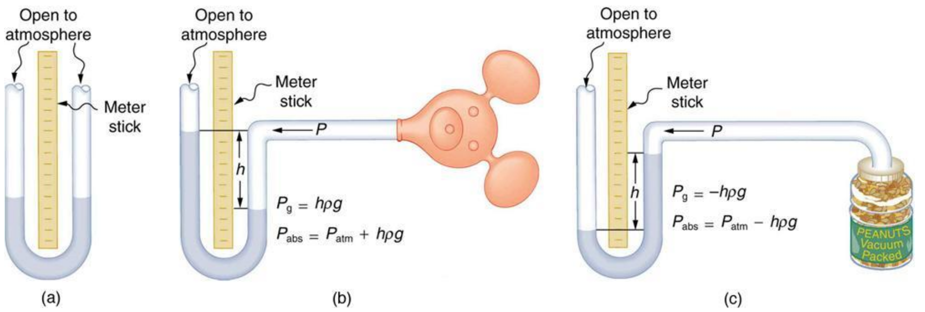


Figure 11.13: An open-tube manometer has one side open to the atmosphere. (a) Fluid depth must be the same on both sides, or the pressure each side exerts at the bottom will be unequal and there will be flow from the deeper side. (b) A positive gauge pressure Pg = $h\rho g$ transmitted to one side of the manometer can support a column of fluid of height h. (c) Similarly, atmospheric pressure is greater than a negative gauge pressure Pg by an amount $h\rho g$. The jar's rigidity prevents atmospheric pressure from being transmitted to the peanuts.

A *barometer* is a device that measures atmospheric pressure. A mercury barometer is shown in Figure 11.14. This device measures atmospheric pressure, rather than gauge pressure, because there is a nearly pure vacuum above the mercury in the tube. The height of the mercury is such that $h\rho g = P_{atm}$. When atmospheric pressure varies, the mercury rises or falls, giving important clues to weather forecasters. The barometer can also be used as an altimeter, since average atmospheric pressure varies with altitude. Mercury barometers and manometers are so common that units of mm Hg are often quoted for atmospheric pressure and blood pressures.

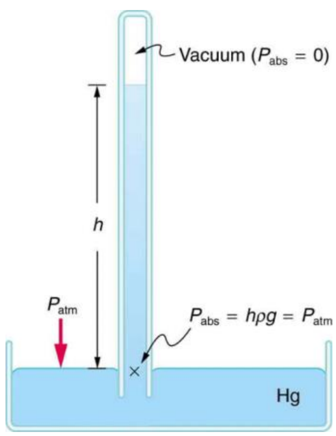


Figure 11.14: A mercury barometer measures atmospheric pressure. The pressure due to the mercury's weight, hρg, equals atmospheric pressure. The atmosphere is able to force mercury in the tube to a height h because the pressure above the mercury is zero.

11.6 Buoyant Force

The buoyant force is the net upward force on any object in any fluid.

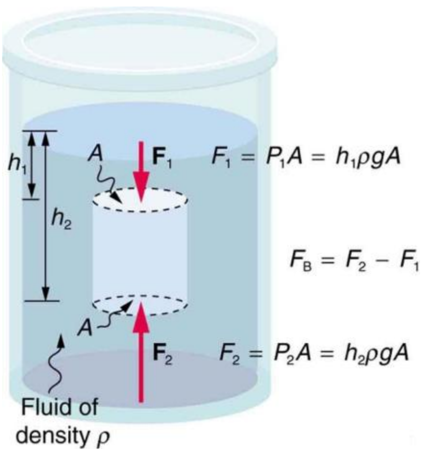


Figure 11.15: Pressure due to the weight of a fluid increases with depth since $P=h\rho g$. This pressure and associated upward force on the bottom of the cylinder are greater than the downward force on the top of the cylinder. Their difference is the buoyant force $\mathbf{F}_B$. (Horizontal forces cancel.)

Just how great is this buoyant force? To answer this question, think about what happens when a submerged object is removed from a fluid, as in Figure 11.16.

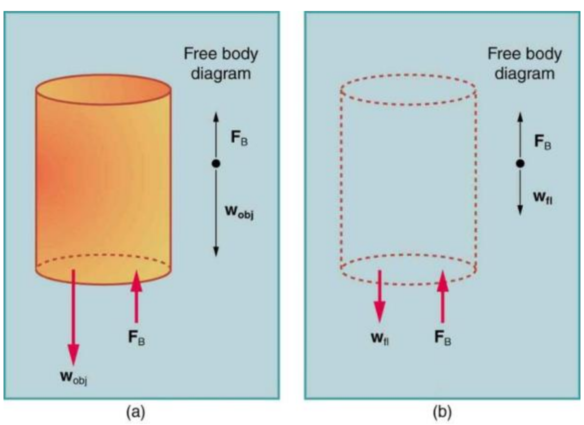


Figure 11.16: (a) An object submerged in a fluid experiences a buoyant force F_B. If F_B is greater than the weight of the object, the object will rise. If F_B is less than the weight of the object, the object will sink. (b) If the object is removed, it is replaced by fluid having weight *w*fl. Since this

weight is supported by surrounding fluid, the buoyant force must equal the weight of the fluid displaced. That is, $F_B = w_{fl}$,a statement of Archimedes' principle.

The space it occupied is filled by fluid having a weight *w*fl. This weight is supported by the surrounding fluid, and so the buoyant force must equal *w*fl, the weight of the fluid displaced by the object. It is a tribute to the genius of the Greek mathematician and inventor Archimedes (ca. 287–212 B.C.) that he stated this principle long before concepts of force were well established. Stated in words, Archimedes' principle is as follows: The buoyant force on an object equals the weight of the fluid it displaces. In equation form,

Archimedes' Principle

According to this principle the buoyant force on an object equals the weight of the fluid it displaces. In equation form, Archimedes' principle is,

$$F_B = w_{fl}$$

where F_B is the buoyant force and *w*fl is the weight of the fluid displaced by the object. Archimedes' principle is valid in general, for any object in any fluid, whether partially or totally submerged.

Floating and Sinking

Drop a lump of clay in water. It will sink. Then mold the lump of clay into the shape of a boat, and it will float. Because of its shape, the boat displaces more water than the lump and experiences a greater buoyant force. The same is true of steel ships.

Example - Calculating buoyant force: dependency on shape

(a) Calculate the buoyant force on 10,000 metric tons (1.00×10^7kg) of solid steel completely submerged in water, and compare this with the steel's weight. (b) What is the maximum buoyant force that water could exert on this same steel if it were shaped into a boat that could displace $1.00\times10^5 m^3$ of water?

Strategy for (a)

To find the buoyant force, we must find the weight of water displaced. We can do this by using the densities of water and steel given in Table 11.1. We note that, since the steel is completely submerged, its volume and the water's volume are the same. Once we know the volume of water, we can find its mass and weight.

Solution for (a)

First, we use the definition of density $\rho = m/V$ to find the steel's volume, and then we substitute values for mass and density. This gives,

$$V_{st} = m_{st} / \rho_{st} = (1.00\times10^7\text{kg}) / (7.8\times10^3\text{kg/m}^3) = 1.28\times10^3\text{m}^3$$

Because the steel is completely submerged, this is also the volume of water displaced, V_w. We can now find the mass of water displaced from the relationship between its volume and density, both of which are known. This gives,

$$m_w = \rho_w V_w = (1.000\times10^3\text{kg/m}^3)(1.28\times10^3\text{m}^3) = 1.28\times10^6\text{kg}$$

By Archimedes' principle, the weight of water displaced is $m_w g$, so the buoyant force is

$$F_B = w_w = m_w g = (1.28\times10^6\text{kg})(9.80\text{m/s}^2) = 1.3\times10^7\text{N}$$

The steel's weight is $m_s g = 9.80\times10^7$N, which is much greater than the buoyant force, so the steel will remain submerged. Note that the buoyant force is rounded to two digits because the density of steel is given to only two digits.

Strategy for (b)

Here we are given the maximum volume of water the steel boat can displace. The buoyant force is the weight of this volume of water.

Solution for (b)

The mass of water displaced is found from its relationship to density and volume, both of which are known. That is,

$$m_w = \rho_w V_w = (1.000\times10^3\text{kg/m}^3)\,(1.00\times10^5\text{m}^3)$$

$$= 1.00\times10^8\text{kg}$$

The maximum buoyant force is the weight of this much water, or

$$F_B = w_w = m_w g = (1.00\times10^8\text{kg})(9.80\text{m/s}^2)$$

$$= 9.80\times10^8\text{N}$$

Discussion

The maximum buoyant force is ten times the weight of the steel, meaning the ship can carry a load nine times its own weight without sinking.

Density and Archimedes' Principle

Density plays a crucial role in Archimedes' principle. The average density of an object is what ultimately determines whether it floats. If its average density is less than that of the surrounding fluid, it will float. This is because the fluid, having a higher density, contains more mass and hence more weight in the same volume. The buoyant force, which equals the weight of the fluid displaced, is thus greater than the weight of the object. Likewise, an object denser than the fluid will sink.

The extent to which a floating object is submerged depends on how the object's density is related to that of the fluid. In Figure 11.17, for example, the unloaded ship has a lower density and less of it is submerged compared with the same ship loaded. We can derive a quantitative expression for the fraction submerged by considering density. The fraction submerged is the ratio of the volume submerged to the volume of the object, or

$$\text{fraction submerged} = V_{\text{sub}} / V_{\text{obj}} = V_{\text{fl}} / V_{\text{obj}}$$

The volume submerged equals the volume of fluid displaced, which we call V_{fl}. Now we can obtain the relationship between the densities by substituting $\rho = m/V$ into the expression. This gives,

$$V_{\text{fl}} / V_{\text{obj}} = (m_{\text{fl}}/\rho_{\text{fl}}) / (m_{\text{obj}}/\rho_{\text{avobj}})$$

where ρ_{avobj} is the average density of the object and ρ_{fl} is the density of the fluid. Since the object floats, its mass and that of the displaced fluid are equal, and so they cancel from the equation, leaving,

$$\text{fraction submerged} = \rho_{\text{avobj}} / \rho_{\text{fl}}$$

Figure 11.17: An unloaded ship (a) floats higher in the water than a loaded ship (b).

We use this last relationship to measure densities. This is done by measuring the fraction of a floating object that is submerged—for example, with a hydrometer. It is useful to define the ratio of the density of an object to a fluid (usually water) as specific gravity:

$$\text{specific gravity} = \rho_{\text{av}} / \rho_{\text{w}}$$

where ρ_{av} is the average density of the object or substance and ρ_w is the density of water at 4.00°C. Specific gravity is dimensionless, independent of whatever units are used for ρ. If an object floats, its specific gravity is less than one. If it sinks, its specific gravity is greater than one. Moreover, the fraction of a floating object that is submerged equals its specific gravity. If an object's specific gravity is exactly 1, then it will remain suspended in the fluid, neither sinking nor floating. Scuba divers try to obtain this state so that they can hover in the water. We measure the specific gravity of fluids, such as battery acid, radiator fluid, and urine, as an indicator of their condition. One device for measuring specific gravity is shown in Figure 11.18.

Specific Gravity

Specific gravity is the ratio of the density of an object to a fluid (usually water).

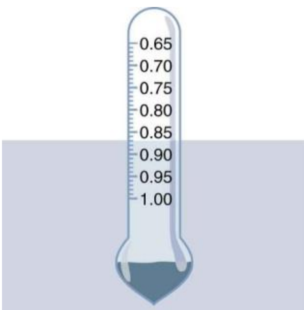


Figure 11.18: This hydrometer is floating in a fluid of specific gravity 0.87. The glass hydrometer is filled with air and weighted with lead at the bottom. It floats highest in the densest fluids and has been calibrated and labeled so that specific gravity can be read from it directly.

More Density Measurements

One of the most common techniques for determining density is shown in Figure 11.19:.

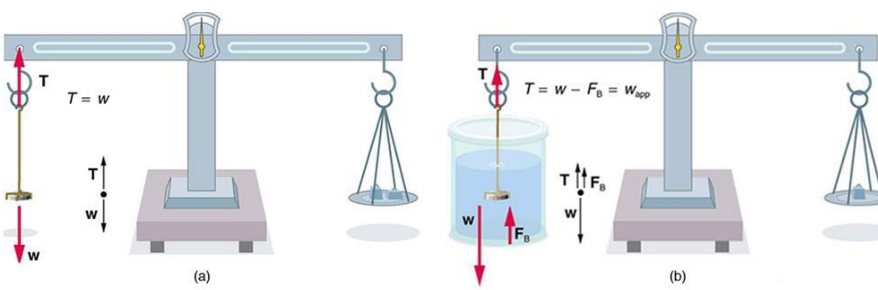


Figure 11.19: (a) A coin is weighed in air. (b) The apparent weight of the coin is determined while it is completely submerged in a fluid of known density. These two measurements are used to calculate the density of the coin.

An object, here a coin, is weighed in air and then weighed again while submerged in a liquid. The density of the coin, an indication of its authenticity, can be calculated if the fluid density is known. This same technique can also be used to determine the density of the fluid if the density of the coin is known. All of these calculations are based on Archimedes' principle.

Archimedes' principle states that the buoyant force on the object equals the weight of the fluid displaced. This, in turn, means that the object *appears* to weigh less when submerged; we call this measurement the object's *apparent weight*. The object suffers an *apparent weight loss* equal to the weight of the fluid displaced. Alternatively, on balances that measure mass, the object suffers an *apparent mass loss* equal to the mass of fluid displaced. That is,

$$\text{apparent weight loss} = \text{weight of fluid displaced}$$

or

$$\text{apparent mass loss} = \text{mass of fluid displaced.}$$

11.7 Cohesion and Adhesion in Liquids

Attractive forces between molecules of the same type are called **cohesive forces**. Liquids can, for example, be held in open containers because cohesive forces hold the molecules together. Attractive forces between molecules of different types are called **adhesive forces**. Such forces cause liquid drops to cling to windowpanes, for example. In this section we examine effects directly attributable to cohesive and adhesive forces in liquids.

Surface Tension

Cohesive forces between molecules cause the surface of a liquid to contract to the smallest possible surface area. This general effect is called surface tension. Molecules on the surface are pulled inward by cohesive forces, reducing the surface area. Molecules inside the liquid experience zero net force, since they have neighbors on all sides.

The model of a liquid surface acting like a stretched elastic sheet can effectively explain surface tension effects. For example, some insects can walk on water (as opposed to floating in it) as we would walk on a trampoline—they dent the surface as shown in Figure 11.20 (a). Figure 11.20 (b) shows another example, where a needle rests on a water surface. The iron needle cannot, and does not, float, because its density is greater than that of water. Rather, its weight is supported by forces in the stretched surface that try to make the surface smaller or flatter. If the needle were placed point down on the surface, its weight acting on a smaller area would break the surface, and it would sink.

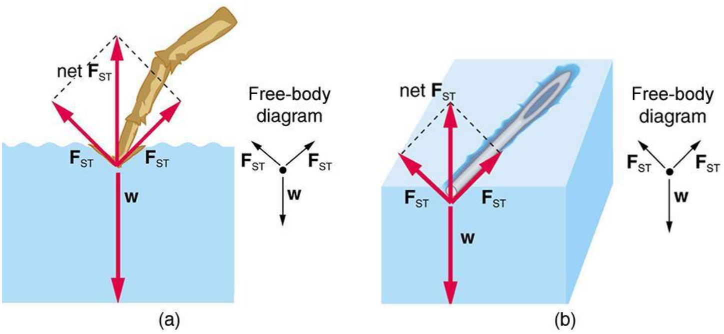


Figure 11.20: Surface tension supporting the weight of an insect and an iron needle, both of which rest on the surface without penetrating it. They are not floating; rather, they are supported by the surface of the liquid. (a) An insect leg dents the water surface. F_{ST} is a restoring force (surface tension) parallel to the surface. (b) An iron needle similarly dents a water surface until the restoring force (surface tension) grows to equal its weight.

Surface tension is proportional to the strength of the cohesive force, which varies with the type of liquid. Surface tension γ is defined to be the force F per unit length L exerted by a stretched liquid membrane:

$$\gamma = F/L$$

For the insect of Figure 11.20 (a), its weight w is supported by the upward components of the surface tension force: $w = \gamma L \sin\theta$, where L is the circumference of the insect's foot in contact with the water. Figure 11.21 shows one way to measure surface tension. The liquid film exerts a force on the movable wire in an attempt to reduce its surface area. The magnitude of this force depends on the surface tension of the liquid and can be measured accurately.

Surface tension is the reason why liquids form bubbles and droplets. The inward surface tension force causes bubbles to be approximately spherical and raises the pressure of the gas trapped inside relative to atmospheric pressure outside. It can be shown that the gauge pressure P inside a spherical bubble is given by,

$$P = 4\gamma/r$$

where r is the radius of the bubble. Thus, the pressure inside a bubble is greatest when the bubble is the smallest. Another bit of evidence for this is illustrated in Figure 11.22. When air is allowed to flow between two balloons of unequal size, the smaller balloon tends to collapse, filling the larger balloon.

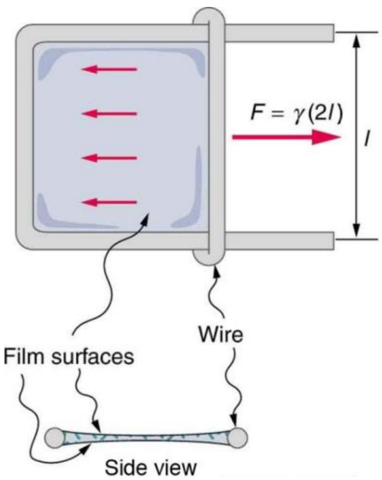


Figure 11.21: Sliding wire device used for measuring surface tension; the device exerts a force to reduce the film's surface area. The force needed to hold the wire in place is $F=\gamma L=\gamma(2l)$, since there are *two* liquid surfaces attached to the wire. This force remains nearly constant as the film is stretched, until the film approaches its breaking point.

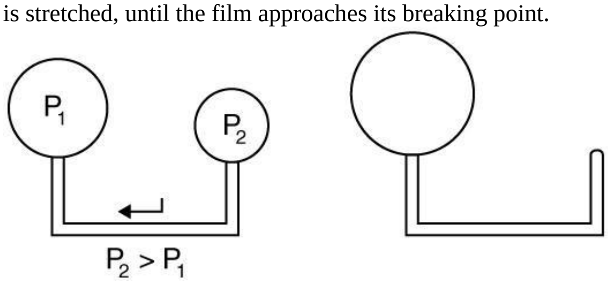


Figure 11.22: With the valve closed, two balloons of different sizes are attached to each end of a tube. Upon opening the valve, the smaller balloon decreases in size with the air moving to fill the larger balloon. The pressure in a spherical balloon is inversely proportional to its radius, so that the smaller balloon has a greater internal pressure than the larger balloon, resulting in this flow.

Adhesion and Capillary Action

Why is it that water beads up on a waxed car but does not on bare paint? The answer is that the adhesive forces between water and wax are much smaller than those between water and paint. Competition between the forces of adhesion and cohesion are important in the macroscopic behavior of liquids. An important factor in studying the roles of these two forces is the angle θ between the tangent to the liquid surface and the surface. (See Figure 11.23.) The contact angle θ is directly related to the relative strength of the cohesive and adhesive forces. The larger the strength of the cohesive force relative to the adhesive force, the larger θ is, and the more the liquid tends to form a droplet. The smaller θ is, the smaller the relative strength, so that the adhesive force is able to flatten the drop.

Contact Angle

The angle θ between the tangent to the liquid surface and the surface is called the contact angle.

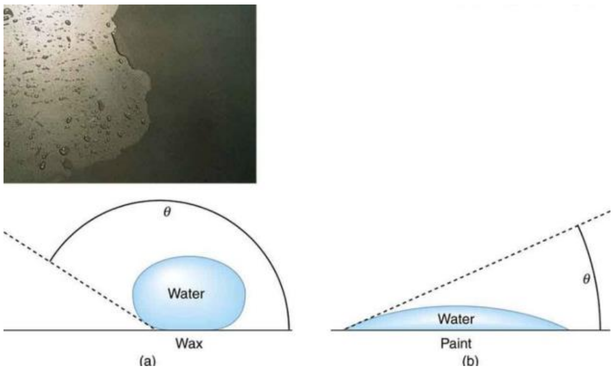


Figure 11.23: In the photograph, water beads on the waxed car paint and flattens on the unwaxed paint. (a) Water forms beads on the waxed surface because the cohesive forces responsible for surface tension are larger than the adhesive forces, which tend to flatten the drop. (b) Water beads on bare paint are flattened considerably because the adhesive forces between water and paint are strong, overcoming surface tension. The contact angle θ is directly related to the relative strengths of the cohesive and adhesive forces. The larger θ is, the larger the ratio of cohesive to adhesive forces. (credit: P. P. Urone)

One important phenomenon related to the relative strength of cohesive and adhesive forces is capillary action—the tendency of a fluid to be raised or suppressed in a narrow tube, or *capillary tube*. This action causes blood to be drawn into a small-diameter tube when the tube touches a drop.

Capillary Action

The tendency of a fluid to be raised or suppressed in a narrow tube, or capillary tube, is called capillary action.

If a capillary tube is placed vertically into a liquid, as shown in Figure 11.24, capillary action will raise or suppress the liquid inside the tube depending on the combination of substances. The actual effect depends on the relative strength of the cohesive and adhesive forces and, thus, the contact angle θ given in the table. If θ is less than 90º, then the fluid will be raised; if θ is greater than 90º, it will be suppressed. Mercury, for example, has a very large surface tension and a large contact angle with glass. When placed in a tube, the surface of a column of mercury curves downward, somewhat like a drop. The curved surface of a fluid in a tube is called a **meniscus**. The tendency of surface tension is always to reduce the surface area. Surface tension thus flattens the curved liquid surface in a capillary tube. This results in a downward force in mercury and an upward force in water, as seen in Figure 11.24.

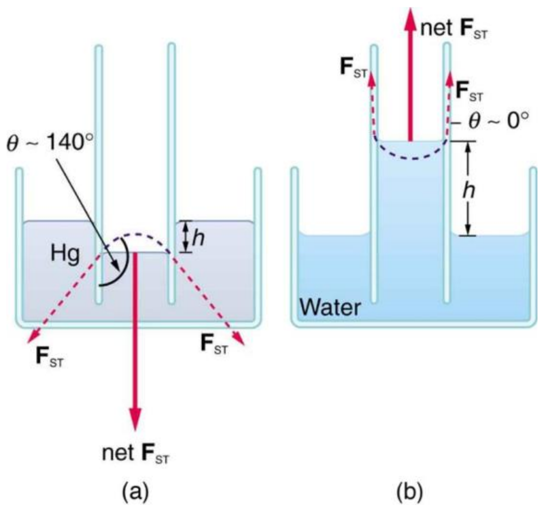


Figure 11.24: (a) Mercury is suppressed in a glass tube because its contact angle is greater than 90º. Surface tension exerts a downward force as it flattens the mercury, suppressing it in the tube. The dashed line shows the shape the mercury surface would have without the flattening effect of surface tension. (b) Water is raised in a glass tube because its contact angle is nearly 0º. Surface tension therefore exerts an upward force when it flattens the surface to reduce its area.

Capillary action can move liquids horizontally over very large distances, but the height to which it can raise or suppress a liquid in a tube is limited by its weight. It can be shown that this height h is given by,

$$h = (2\gamma\cos\theta) / (\rho g r)$$

If we look at the different factors in this expression, we might see how it makes good sense. The height is directly proportional to the surface tension γ, which is its direct cause. Furthermore, the height is inversely proportional to tube radius—the smaller the radius r, the higher the fluid can be raised, since a smaller tube holds less mass. The height is also inversely proportional to fluid density ρ, since a larger density means a greater mass in the same volume. (See Figure 11.25)

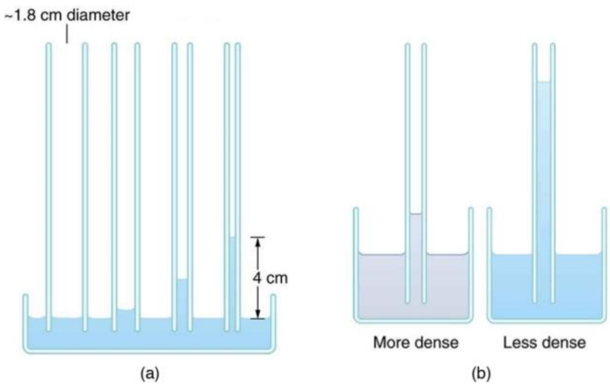


Figure 11.25: (a) Capillary action depends on the radius of a tube. The smaller the tube, the greater the height reached. The height is negligible for large-radius tubes. (b) A denser fluid in the same tube rises to a smaller height, all other factors being the same.

For examples and answers, please refer to OpenStax.com questions and answers given on their website or to the College Physics 2e - https://openstax.org/details/books/college-physics-2e.

Chapter 12

12.0 Objectives

At the end of this lesson, students should be able to,

- Apply Bernoulli's equation.
- Determine flow rate and corresponding velocities.
- Apply Poiseuille's law.
- Identify flow as laminar or turbulent through application of Reynold's number.
- Identify osmosis, diffusion, and active transport.

12.1 Introduction

This chapter covers Bernoulli's equation, Poiseuille's law, flow rate and velocities, Reynold's number, osmosis, diffusion, and active transport, and related calculations.

We have dealt with many situations in which fluids are static. But by their very definition, fluids flow. Examples come easily—a column of smoke rises from a campfire, water streams from a fire hose, blood courses through your veins. Why does rising smoke curl and twist? How does a nozzle increase the speed of water emerging from a hose? How does the body regulate blood flow? The physics of fluids in motion—fluid dynamics—allows us to answer these and many other questions.

12.2 Flow Rate

Flow rate *Q* is defined to be the volume of fluid passing by some location through an area during a period of time, as seen in Figure 12.1. In symbols, this can be written as,

$$Q = V/t$$

where *V* is the volume and *t* is the elapsed time.

The SI unit for flow rate is m^3/s, but a number of other units for *Q* are in common use. For example, the heart of a resting adult pumps blood at a rate of 5.00 liters per minute (L/min). Note that a liter (L) is 1/1000 of a cubic meter or 1000 cubic centimeters ($10^{-3} m^3$ or $10^3 cm^3$). In this text we shall use whatever metric units are most convenient for a given situation.

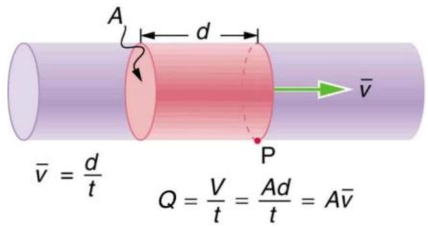


Figure 12.1: Flow rate is the volume of fluid per unit time flowing past a point through area *A*. Here the shaded cylinder of fluid flows past point P in a uniform pipe in time *t*. The volume of the cylinder is *Ad* and the average velocity is $v_{av} = d/t$ so that the flow rate is $Q = Ad/t = Av_{av}$.

Example - Calculating Volume from Flow Rate: The Heart Pumps a Lot of Blood in a Lifetime

How many cubic meters of blood does the heart pump in a 75-year lifetime, assuming the average flow rate is 5.00 L/min?

Strategy

Time and flow rate *Q* are given, and so the volume *V* can be calculated from the definition of flow rate.

Solution

Solving $Q = V/t$ for volume gives

$$V = Qt.$$

Substituting known values yields

$$V = (5.00\text{L}/1\text{ min})\,(75\text{y})\,(1\text{m}^3/10^3\text{L})(5.26\times10^5\text{ min/y})$$

$$= 2.0\times10^5\text{m}^3$$

Discussion

This amount is about 200,000 tons of blood. For comparison, this value is equivalent to about 200 times the volume of water contained in a 6-lane 50-m lap pool.

Flow rate and velocity are related, but quite different, physical quantities. To make the distinction clear, think about the flow rate of a river. The greater the velocity of the water, the greater the flow rate of the river. But flow rate also depends on the size of the river. A rapid mountain stream carries far less water than the Amazon River in Brazil, for example. The precise relationship between flow rate *Q* and velocity v_{av} is,

$$Q = Av_{av}$$

where A is the cross-sectional area and v_{av} is the average velocity. This equation seems logical enough. The relationship tells us that flow rate is directly proportional to both the magnitude of the average velocity (hereafter referred to as the speed) and the size of a river, pipe, or other conduit. The larger the conduit, the greater its cross-sectional area. Figure 12.1 illustrates how this relationship is obtained. The shaded cylinder has a volume,

$$V = Ad$$

which flows past the point P in a time t. Dividing both sides of this relationship by t gives,

$$V/t = Ad/t$$

We note that $Q = V/t$ and the average speed is $v_{av} = d/t$. Thus, the equation becomes $Q = Av_{av}$.

Figure 12.2 shows an incompressible fluid flowing along a pipe of decreasing radius. Because the fluid is incompressible, the same amount of fluid must flow past any point in the tube in a given time to ensure continuity of flow. In this case, because the cross-sectional area of the pipe decreases, the velocity must necessarily increase. This logic can be extended to say that the flow rate must be the same at all points along the pipe. In particular, for points 1 and 2,

$$Q_1 = Q_2$$

$$A_1 v_{av1} = A_2 v_{av2}$$

This is called the equation of continuity and is valid for any incompressible fluid. The consequences of the equation of continuity can be observed when water flows from a hose into a narrow spray nozzle: it emerges with a large speed—that is the purpose of the nozzle. Conversely, when a river empties into one end of a reservoir, the water slows considerably, perhaps picking up speed again when it leaves the other end of the reservoir. In other words, speed increases when cross-sectional area decreases, and speed decreases when cross-sectional area increases.

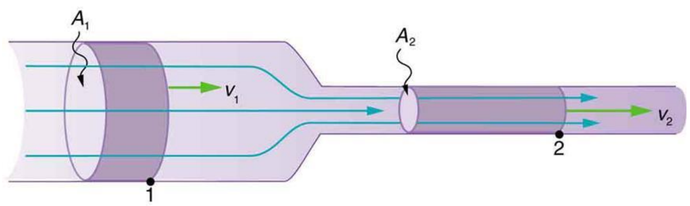


Figure 12.2: When a tube narrows, the same volume occupies a greater length. For the same volume to pass points 1 and 2 in a given time, the speed must be greater at point 2. The process is reversible. If the fluid flows in the opposite direction, its speed will decrease when the tube widens. (Note that the relative volumes of the two cylinders and the corresponding velocity vector arrows are not drawn to scale.)

Since liquids are essentially incompressible, the equation of continuity is valid for all liquids. However, gases are compressible, and so the equation must be applied with caution to gases if they are subjected to compression or expansion.

Example - Calculating Fluid Speed: Speed Increases When a Tube Narrows

A nozzle with a radius of 0.250 cm is attached to a garden hose with a radius of 0.900 cm. The flow rate through hose and nozzle is 0.500 L/s. Calculate the speed of the water (a) in the hose and (b) in the nozzle.

Strategy

We can use the relationship between flow rate and speed to find both velocities. We will use the subscript 1 for the hose and 2 for the nozzle.

Solution for (a)

First, we solve $Q = Av_{av}$ for v_1 and note that the cross-sectional area is $A = \pi r^2$, yielding

$$v_{av1} = Q/A_1 = Q/\pi r^2{}_1$$

Substituting known values and making appropriate unit conversions yields

$$v_{av1} = [(0.500\text{L/s})\ (10^{-3}\text{m}^3/\text{L})] \ / \ [\pi(9.00\times10^{-3}\text{m})^2] = 1.96 \text{ m/s}.$$

Solution for (b)

We could repeat this calculation to find the speed in the nozzle $\bar{v}2$, but we will use the equation of continuity to give a somewhat different insight. Using the equation which states

$$A_1 v_{av1} = A_2 v_{av2}$$

solving for v_{av2} and substituting πr^2 for the cross-sectional area yields

$$v_{av2} = (A_1 / A_2)v_{av1} = (\pi r^2{}_1 / \pi r^2{}_2)v_{av1} = (r_1{}^2 / r_2{}^2)\ v_{av1}$$

Substituting known values,

$$v_{av2} = [(0.900\text{cm})^2(0.250\text{cm})^2]1.96 \text{ m/s} = 25.5 \text{ m/s}.$$

Discussion

A speed of 1.96 m/s is about right for water emerging from a nozzleless hose. The nozzle produces a considerably faster stream merely by constricting the flow to a narrower tube.

The solution to the last part of the example shows that speed is inversely proportional to the *square* of the radius of the tube, making for large effects when radius varies. We can blow out a candle at quite a distance, for example, by pursing our lips, whereas blowing on a candle with our mouth wide open is quite ineffective.

In many situations, including in the cardiovascular system, branching of the flow occurs. The blood is pumped from the heart into arteries that subdivide into smaller arteries (arterioles) which branch into very fine vessels called capillaries. In this situation, continuity of flow is maintained but it is the *sum* of the flow rates in each of the branches in any portion along the tube that is maintained. The equation of continuity in a more general form becomes

$$n_1 A_1 v_{\mathrm{av1}} = n_2 A_2 v_{\mathrm{av2}}$$

where n_1 and n_2 are the number of branches in each of the sections along the tube.

Example - Calculating Flow Speed and Vessel Diameter: Branching in the Cardiovascular System

The aorta is the principal blood vessel through which blood leaves the heart in order to circulate around the body. (a) Calculate the average speed of the blood in the aorta if the flow rate is 5.0 L/min. The aorta has a radius of 10 mm. (b) Blood also flows through smaller blood vessels known as capillaries. When the rate of blood flow in the aorta is 5.0 L/min, the speed of blood in the capillaries is about 0.33 mm/s. Given that the average diameter of a capillary is 8.0μm, calculate the number of capillaries in the blood circulatory system.

Strategy

We can use $Q = Av_{\mathrm{av}}$ to calculate the speed of flow in the aorta and then use the general form of the equation of continuity to calculate the number of capillaries as all of the other variables are known.

Solution for (a)

The flow rate is given by $Q=Av_{av}$ or $v_{av}=Q\pi r^2$ for a cylindrical vessel.

Substituting the known values (converted to units of meters and seconds) gives,

$$v_{\mathrm{av}} = [(5.0\mathrm{L/min})(10^{-3}\mathrm{m^3/L})(1\mathrm{min}/60\mathrm{s})] / [\pi(0.010\ \mathrm{m})^2] = 0.27\mathrm{m/s}$$

Solution for (b)

Using $n_1 A_1 v_{\mathrm{av1}} = n_2 A_2 v_{\mathrm{av1}}$, assigning the subscript 1 to the aorta and 2 to the capillaries, and solving for n_2 (the number of capillaries) gives $n_2 = n_1 A_1 v_{\mathrm{av1}} / A_2 v_{av2}$. Converting all quantities to units of meters and seconds and substituting into the equation above gives,

$$n_2 = [(1)(\pi)(10\times10^{-3}\mathrm{m})^2(0.27\ \mathrm{m/s})] / [(\pi)(4.0\times10^{-6}\mathrm{m})^2(0.33\times10^{-3}\mathrm{m/s}) = 5\ .0\times10^9 \mathrm{capillaries}.$$

Discussion

Note that the speed of flow in the capillaries is considerably reduced relative to the speed in the aorta due to the significant increase in the total cross-sectional area at the capillaries. This low speed is to allow sufficient time for effective exchange to occur although it is equally important for the flow not to become stationary in order to avoid the possibility of clotting. Does this large number of capillaries in the body seem reasonable? In active muscle, one finds about 200 capillaries per mm3, or about 200×10^6 per 1 kg of muscle. For 20 kg of muscle, this amounts to about 4×10^9 capillaries.

When a fluid flows into a narrower channel, its speed increases. That means its kinetic energy also increases. Where does that change in kinetic energy come from? The increased kinetic energy comes from the net work done on the fluid to push it into the channel and the work done on the fluid by the gravitational force, if the fluid changes vertical position. Recall the work-energy theorem,

$$W_{\mathrm{net}} = \tfrac{1}{2}\, mv^2 - \tfrac{1}{2}\, mv^2{}_0$$

There is a pressure difference when the channel narrows. This pressure difference results in a net force on the fluid: recall that pressure times area equals force. The net work done increases the fluid's kinetic energy. As a result, the *pressure will drop in a rapidly-moving fluid*, whether or not the fluid is confined to a tube.

There are a number of common examples of pressure dropping in rapidly-moving fluids. Shower curtains have a disagreeable habit of bulging into the shower stall when the shower is on. The high-velocity stream of water and air creates a region of lower pressure inside the shower, and standard atmospheric pressure on the other side. The pressure difference results in a net force inward pushing the curtain in. You may also have noticed that when passing a truck on the highway, your car tends to veer toward it. The reason is the same—the high velocity of the air between the car and the truck creates a region of lower pressure, and the vehicles are pushed together by greater pressure on the outside. (See Figure 12.3.) This effect was observed as far back as the mid-1800s, when it was found that trains passing in opposite directions tipped precariously toward one another.

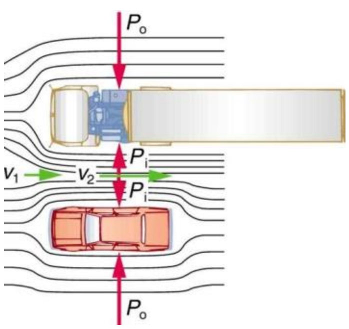


Figure 12.3: An overhead view of a car passing a truck on a highway. Air passing between the vehicles flows in a narrower channel and must increase its speed (v_2 is greater than v_1), causing the pressure between them to drop (P_i is less than P_o). Greater pressure on the outside pushes the car and truck together.

12.3 Bernoulli's Equation

The relationship between pressure and velocity in fluids is described quantitatively by Bernoulli's equation, named after its discoverer, the Swiss scientist Daniel Bernoulli (1700–1782). Bernoulli's equation states that for an incompressible, frictionless fluid, the following sum is constant:

$$P + \frac{1}{2}\rho v^2 + \rho g h = \text{constant}$$

where P is the absolute pressure, ρ is the fluid density, v is the velocity of the fluid, h is the height above some reference point, and g is the acceleration due to gravity. If we follow a small volume of fluid along its path, various quantities in the sum may change, but the total remains constant. Let the subscripts 1 and 2 refer to any two points along the path that the bit of fluid follows; Bernoulli's equation becomes,

$$P_1 + \frac{1}{2}\rho v^2{}_1 + \rho g h_1 = P_2 + \frac{1}{2}\rho v_2{}^2 + \rho g h_2$$

Bernoulli's equation is a form of the conservation of energy principle. Note that the second and third terms are the kinetic and potential energy with m replaced by ρ. In fact, each term in the equation has units of energy per unit volume. We can prove this for the second term by substituting $\rho = m/V$ into it and gathering terms:

$$\frac{1}{2}\rho v^2 = \frac{1}{2} m v^2 / V = \mathrm{KE}/V$$

So 12ρv2 is the kinetic energy per unit volume. Making the same substitution into the third term in the equation, we find

$$Pgh = mgh/V = \mathrm{PEg}\ /V$$

so ρgh is the gravitational potential energy per unit volume. Note that pressure P has units of energy per unit volume, too. Since $P=F/A$, its units are $\mathrm{N/m^2}$. If we multiply these by m/m, we obtain $\mathrm{N{\cdot}m/m^3} = \mathrm{J/m^3}$, or energy per unit volume. Bernoulli's equation is, in fact, just a convenient statement of conservation of energy for an incompressible fluid in the absence of friction.

The general form of Bernoulli's equation has three terms in it, and it is broadly applicable. To understand it better, we will look at a number of specific situations that simplify and illustrate its use and meaning.

Bernoulli's Equation for Static Fluids

Let us first consider the very simple situation where the fluid is static — that is, $v_1 = v_2 = 0$. Bernoulli's equation in that case is,

$$P_1 + \rho g h_1 = P_2 + \rho g h_2$$

We can further simplify the equation by taking $h2=0$ (we can always choose some height to be zero, just as we often have done for other situations involving the gravitational force, and take all other heights to be relative to this). In that case, we get

$$P_2 = P_1 + \rho g h_1$$

This equation tells us that, in static fluids, pressure increases with depth. As we go from point 1 to point 2 in the fluid, the depth increases by $h1$, and consequently, $P2$ is greater than $P1$ by an amount $\rho gh1$. In the very simplest case, $P1$ is zero at the top of the fluid, and we get the familiar relationship $P = \rho gh$. (Recall that $P = \rho gh$ and $\Delta \text{PEg} = mgh$.) Bernoulli's equation includes the fact that the pressure due to the weight of a fluid is ρgh.

Bernoulli's Principle—Bernoulli's Equation at Constant Depth

Another important situation is one in which the fluid moves but its depth is constant - that is, $h_1=h_2$. Under that condition, Bernoulli's equation becomes,

$$P_1 + ½\, \rho v^2{}_1 = P_2 + ½\, \rho v^2{}_2$$

Situations in which fluid flows at a constant depth are so important that this equation is often called Bernoulli's principle. It is Bernoulli's equation for fluids at constant depth. (Note again that this applies to a small volume of fluid as we follow it along its path.) As we have just discussed, pressure drops as speed increases in a moving fluid. We can see this from Bernoulli's principle. For example, if v_2 is greater than v_1 in the equation, then P_2 must be less than P_1 for the equality to hold.

Applications of Bernoulli's Principle

There are a number of devices and situations in which fluid flows at a constant height and, thus, can be analyzed with Bernoulli's principle.

Entrainment

People have long put the Bernoulli principle to work by using reduced pressure in high-velocity fluids to move things about. With a higher pressure on the outside, the high-velocity fluid forces other fluids into the stream. This process is called *entrainment*. Entrainment devices have been in use since ancient times, particularly as pumps to raise water small heights, as in draining swamps, fields, or other low-lying areas. Some other devices that use the concept of entrainment are shown in Figure 12.4.

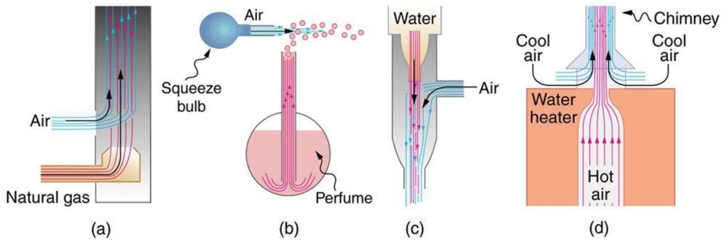


Figure 12.4: Examples of entrainment devices that use increased fluid speed to create low pressures, which then entrain one fluid into another. (a) A Bunsen burner uses an adjustable gas nozzle, entraining air for proper combustion. (b) An atomizer uses a squeeze bulb to create a jet of air that entrains drops of perfume. Paint sprayers and carburetors use very similar techniques to move their respective liquids. (c) A common aspirator uses a high-speed stream of water to create a region of lower pressure. Aspirators may be used as suction pumps in dental and surgical situations or for draining a flooded basement or producing a reduced pressure in a vessel. (d) The chimney of a water heater is designed to entrain air into the pipe leading through the ceiling.

Wings and Sails

The airplane wing is a beautiful example of Bernoulli's principle in action. Figure 12.5 (a) shows the characteristic shape of a wing. The wing is tilted upward at a small angle and the upper surface is longer, causing air to flow faster over it. The pressure on top of the wing is therefore reduced, creating a net upward force or lift. (Wings can also gain lift by pushing air downward, utilizing the conservation of momentum principle. The deflected air molecules result in an upward force on the wing — Newton's third law.) Sails also have the characteristic shape of a wing. (See Figure 12.5 (b).) The pressure on the front side of the sail, P_{front}, is lower than the pressure on the back of the sail, P_{back}. This results in a forward force and even allows you to sail into the wind.

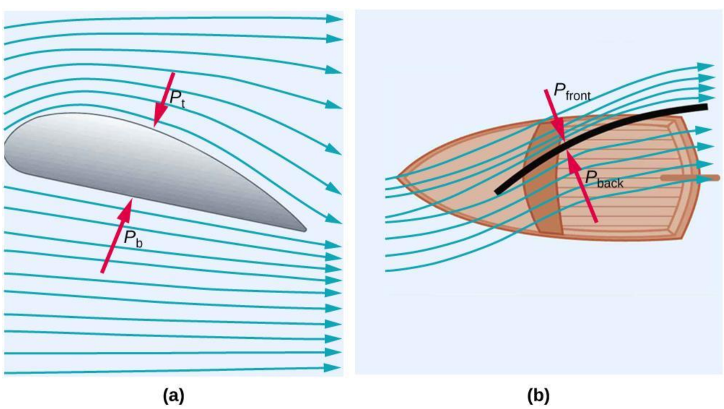


Figure 12.5: (a) The Bernoulli principle helps explain lift generated by a wing. (b) Sails use the same technique to generate part of their thrust.

Velocity measurement

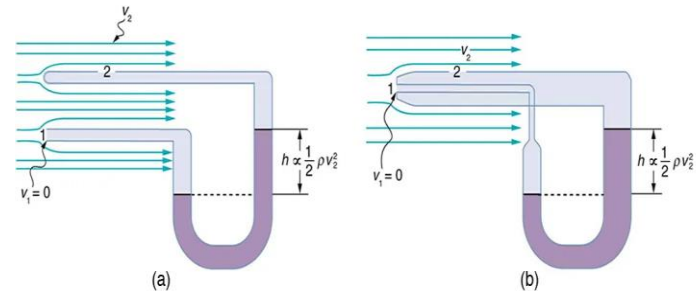


Figure 12.6: Measurement of fluid speed based on Bernoulli's principle. (a) A manometer is connected to two tubes that are close together and small enough not to disturb the flow. Tube 1 is open at the end facing the flow. A dead spot having zero speed is created there. Tube 2 has an opening on the side, and so the fluid has a speed v across the opening; thus, pressure there drops. The difference in pressure at the manometer is ½ ρv_2^2, and so h is proportional to ½ ρv_2^2. (b) This type of velocity measuring device is a Prandtl tube, also known as a pitot tube.

Figure 12.6 shows two devices that measure fluid velocity based on Bernoulli's principle. The manometer in Figure 12.6 (a) is connected to two tubes that are small enough not to appreciably disturb the flow. The tube facing the oncoming fluid creates a dead spot having zero velocity ($v1=0$) in front of it, while fluid passing the other tube has velocity $v2$. This means that Bernoulli's principle as stated in $P_1+ ½\, \rho v^2{}_1= P_2+ ½\, \rho v^2{}_2$ becomes

$$P_1 = P_2 + ½\, \rho v^2{}_2$$

Thus pressure P_2 over the second opening is reduced by ½ ρv_2^2, and so the fluid in the manometer rises by h on the side connected to the second opening, where

$$h \propto ½\, \rho v_2^2$$

(Recall that the symbol ∝ means "proportional to.") Solving for $v2$, we see that,

$$v_2 \propto \sqrt{h}$$

Figure 12.6 (b) shows a version of this device that is in common use for measuring various fluid velocities; such devices are frequently used as air speed indicators in aircraft.

12.4 Torricelli's Theorem

Figure 12.7 shows water gushing from a large tube through a dam. What is its speed as it emerges? Interestingly, if resistance is negligible, the speed is just what it would be if the water fell a distance *h* from the surface of the reservoir; the water's speed is independent of the size of the opening. Let us check this out. Bernoulli's equation must be used since the depth is not constant. We consider water flowing from the surface (point 1) to the tube's outlet (point 2). Bernoulli's equation as stated in previously is

$$P_1 + \frac{1}{2}\rho v^2{}_1 + \rho g h_1 = P_2 + \frac{1}{2}\rho v_2{}^2 + \rho g h_2$$

Both P_1 and P_2 equal atmospheric pressure (P_1 is atmospheric pressure because it is the pressure at the top of the reservoir. P_2 must be atmospheric pressure, since the emerging water is surrounded by the atmosphere and cannot have a pressure different from atmospheric pressure.) and subtract out of the equation, leaving,

$$\frac{1}{2}\rho v^2{}_1 + \rho g h_1 = \frac{1}{2}\rho v_2{}^2 + \rho g h_2$$

Solving this equation for $v_2{}^2$, noting that the density ρ cancels (because the fluid is incompressible), yields,

$$v_2{}^2 = v_2{}^1 + 2g(h_1 - h_2)$$

We let $h = h_1 - h_2$; the equation then becomes

$$v_2{}^2 = v_2{}^1 + 2gh$$

where *h* is the height dropped by the water. This is simply a kinematic equation for any object falling a distance *h* with negligible resistance. In fluids, this last equation is called *Torricelli's theorem*. Note that the result is independent of the velocity's direction, just as we found when applying conservation of energy to falling objects.

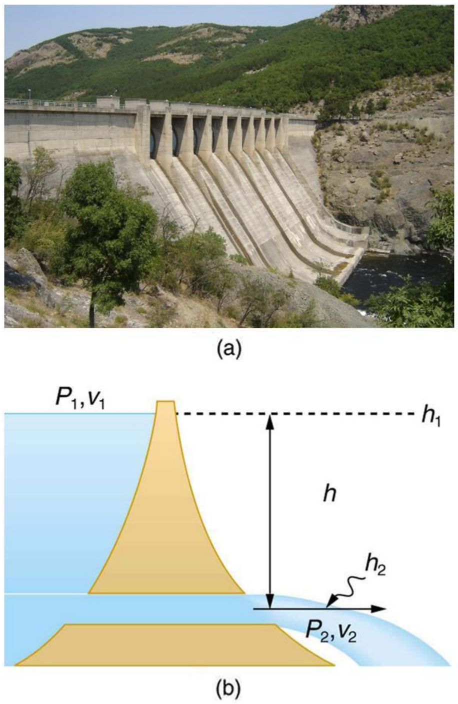


Figure 12.7: (a) Water gushes from the base of the Studen Kladenetz dam in Bulgaria. (credit: Kiril Kapustin; http://www.ImagesFromBulgaria.com) (b) In the absence of significant resistance, water flows from the reservoir with the same speed it would have if it fell the distance *h* without friction. This is an example of Torricelli's theorem.

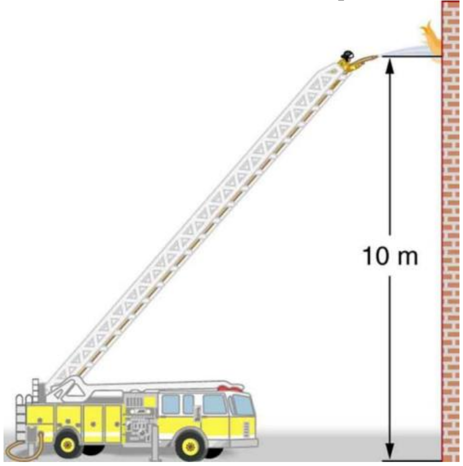


Figure 12.8: Pressure in the nozzle of this fire hose is less than at ground level for two reasons: the water has to go uphill to get to the nozzle, and speed increases in the nozzle. In spite of its lowered pressure, the water can exert a large force on anything it strikes, by virtue of its kinetic energy. Pressure in the water stream becomes equal to atmospheric pressure once it emerges into the air.

Power in Fluid Flow

Power is the *rate* at which work is done or energy in any form is used or supplied. To see the relationship of power to fluid flow, consider Bernoulli's equation:

$$P + \tfrac{1}{2}\rho v^2 + \rho g h = \text{constant}$$

All three terms have units of energy per unit volume, as discussed in the previous section. Now, considering units, if we multiply energy per unit volume by flow rate (volume per unit time), we get units of power. That is, $(E/V)(V/t)=E/t$. This means that if we multiply Bernoulli's equation by flow rate Q, we get power. In equation form, this is,

$$(P + \tfrac{1}{2}\rho v^2+\rho g h)\, Q = \text{power.}$$

Each term has a clear physical meaning. For example, PQ is the power supplied to a fluid, perhaps by a pump, to give it its pressure P. Similarly, $\tfrac{1}{2}\rho v^2 Q$ is the power supplied to a fluid to give it its kinetic energy. And $\rho g h Q$ is the power going to gravitational potential energy.

12.5 Laminar Flow and Viscosity

When you pour yourself a glass of juice, the liquid flows freely and quickly. But when you pour syrup on your pancakes, that liquid flows slowly and sticks to the pitcher. The difference is fluid friction, both within the fluid itself and between the fluid and its surroundings. We call this property of fluids *viscosity*. Juice has low viscosity, whereas syrup has high viscosity. In the previous sections we have considered ideal fluids with little or no viscosity. In this section, we will investigate what factors, including viscosity, affect the rate of fluid flow.

The precise definition of viscosity is based on *laminar*, or nonturbulent, flow. Before we can define viscosity, then, we need to define laminar flow and turbulent flow. Figure 12.9 shows both types of flow. Laminar flow is characterized by the smooth flow of the fluid in layers that do not mix. Turbulent flow, or turbulence, is characterized by eddies and swirls that mix layers of fluid together.

Figure 12.9: Smoke rises smoothly for a while and then begins to form swirls and eddies. The smooth flow is called laminar flow, whereas the swirls and eddies typify turbulent flow. If you watch the smoke (being careful not to breathe on it), you will notice that it rises more rapidly when flowing smoothly than after it becomes turbulent, implying that turbulence poses more resistance to flow. (credit: Creativity103)

Figure 12.10 shows schematically how laminar and turbulent flow differ. Layers flow without mixing when flow is laminar. When there is turbulence, the layers mix, and there are significant velocities in directions other than the overall direction of flow. The lines that are shown in many illustrations are the paths followed by small volumes of fluids. These are called *streamlines*. Streamlines are smooth and continuous when flow is laminar, but break up and mix when flow is turbulent. Turbulence has two main causes. First, any obstruction or sharp corner, such as in a faucet, creates turbulence by imparting velocities perpendicular to the flow. Second, high speeds cause turbulence. The drag both between adjacent layers of fluid and between the fluid and its surroundings forms swirls and eddies, if the speed is great enough.

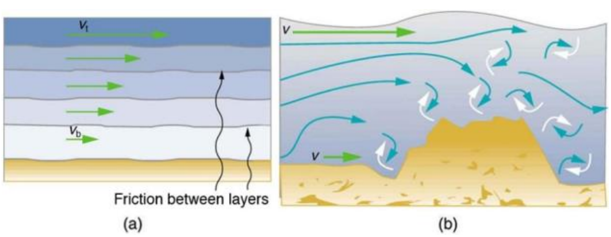


Figure 12.10 (a): Laminar flow occurs in layers without mixing. Notice that viscosity causes drag between layers as well as with the fixed surface. **(b)** An obstruction in the vessel produces turbulence. Turbulent flow mixes the fluid. There is more interaction, greater heating, and more resistance than in laminar flow.

Figure 12.11 shows how viscosity is measured for a fluid. Two parallel plates have the specific fluid between them. The bottom plate is held fixed, while the top plate is moved to the right, dragging fluid with it. The layer (or lamina) of fluid in contact with either plate does not move relative to the plate, and so the top layer moves at v while the bottom layer remains at rest. Each successive layer from the top down exerts a force on the one below it, trying to drag it along, producing a continuous variation in speed from v to 0 as shown. Care is taken to insure that the flow is laminar; that is, the layers do not mix. The motion in Figure 12.11 is like a continuous shearing motion. Fluids have zero shear strength, but the *rate* at which they are sheared is related to the same geometrical factors A and L as is shear deformation for solids.

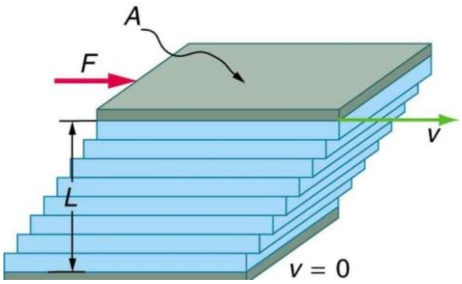


Figure 12.11: The graphic shows laminar flow of fluid between two plates of area A. The bottom plate is fixed. When the top plate is pushed to the right, it drags the fluid along with it.

A force F is required to keep the top plate in Figure 12.11 moving at a constant velocity v, and experiments have shown that this force depends on four factors. First, F is directly proportional to v (until the speed is so high that turbulence occurs—then a much larger force is needed, and it has a more complicated dependence on v). Second, F is proportional to the area A of the plate. This relationship seems reasonable, since A is directly proportional to the amount of fluid being moved. Third, F is inversely proportional to the distance between the plates L. This relationship is also reasonable; L is like a lever arm, and the greater the lever arm, the less force that is needed. Fourth, F is directly proportional to *the coefficient of viscosity*, η. The greater the viscosity, the greater the force required. These dependencies are combined into the equation,

$$F = \eta\ (vA/L)$$

which gives us a working definition of fluid viscosity η. Solving for η gives,

$$\eta = (FL/vA)$$

which defines viscosity in terms of how it is measured. The SI unit of viscosity is $\mathrm{N{\cdot}m/[(m/s)m^2]=(N/m^2)s}$ or $\mathrm{Pa{\cdot}s}$.

Viscosity varies from one fluid to another by several orders of magnitude. As you might expect, the viscosities of gases are much less than those of liquids, and these viscosities are often temperature dependent. The viscosity of blood can be reduced by aspirin consumption, allowing it to flow more easily around the body. (When used over the long term in low doses, aspirin can help prevent heart attacks, and reduce the risk of blood clotting.)

12.6 Laminar Flow Confined to Tubes—Poiseuille's Law

What causes flow? The answer, not surprisingly, is pressure difference. In fact, there is a very simple relationship between horizontal flow and pressure. Flow rate Q is in the direction from high to low pressure. The greater the pressure differential between two points, the greater the flow rate. This relationship can be stated as

$$Q = (P_2 - P_1 / R)$$

where P_1 and P_2 are the pressures at two points, such as at either end of a tube, and R is the resistance to flow. The resistance R includes everything, except pressure that affects flow rate. For example, R is greater for a long tube than for a short one. The greater the viscosity of a fluid, the greater the value of R. Turbulence greatly increases R, whereas increasing the diameter of a tube decreases R.

If viscosity is zero, the fluid is frictionless and the resistance to flow is also zero. Comparing frictionless flow in a tube to viscous flow, as in Figure 12.12, we see that for a viscous fluid, speed is greatest at midstream because of drag at the boundaries. We can see the effect of viscosity in a Bunsen burner flame, even though the viscosity of natural gas is small.

The resistance R to laminar flow of an incompressible fluid having viscosity η through a horizontal tube of uniform radius r and length l, such as the one in Figure 12.13, is given by

$$R = (8\eta l) / (\pi r 4)$$

This equation is called Poiseuille's law for resistance after the French scientist J. L. Poiseuille (1799–1869), who derived it in an attempt to understand the flow of blood, an often turbulent fluid.

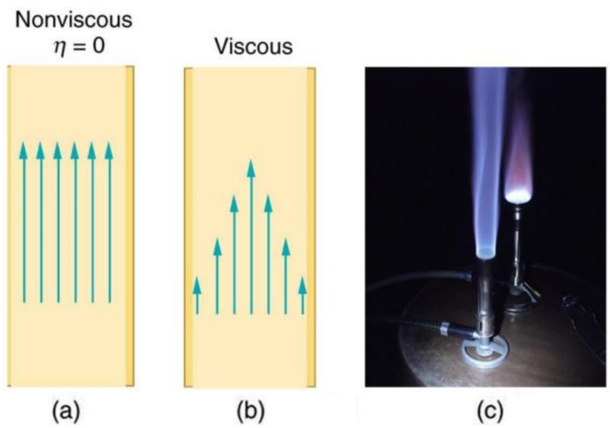


Figure 12.12: (a) If fluid flow in a tube has negligible resistance, the speed is the same all across the tube. (b) When a viscous fluid flows through a tube, its speed at the walls is zero, increasing steadily to its maximum at the center of the tube. (c) The shape of the Bunsen burner flame is due to the velocity profile across the tube. (credit: Jason Woodhead)

Let us examine Poiseuille's expression for R to see if it makes good intuitive sense. We see that resistance is directly proportional to both fluid viscosity η and the length l of a tube. After all, both of these directly affect the amount of friction encountered—the greater either is, the greater the resistance and the smaller the flow. The radius r of a tube affects the resistance, which again makes sense, because the greater the radius, the greater the flow (all other factors remaining the same). But it is surprising that r is raised to the *fourth* power in Poiseuille's law. This exponent means that any change in the radius of a tube has a very large effect on resistance. For example, doubling the radius of a tube decreases resistance by a factor of 24=16.

Taken together, $Q = (P_2 - P\,1)/R$ and $R = (8\eta l)/(\pi r^4)$ give the following expression for flow rate:

$$Q = [(P2 - P1)\pi r^4\,] / (8\eta l)$$

This equation describes laminar flow through a tube. It is sometimes called Poiseuille's law for laminar flow, or simply Poiseuille's law.

The circulatory system provides many examples of Poiseuille's law in action—with blood flow regulated by changes in vessel size and blood pressure. Blood vessels are not rigid but elastic. Adjustments to blood flow are primarily made by varying the size of the vessels, since the resistance is so sensitive to the radius. During vigorous exercise, blood vessels are selectively dilated to important muscles and organs and blood pressure increases. This creates both greater overall blood flow and increased flow to specific areas. Conversely, decreases in vessel radii, perhaps from plaques in the arteries, can greatly reduce blood flow. If a vessel's radius is reduced by only 5% (to 0.95 of its original value), the flow rate is reduced to about (0.95)4 = 0.81 of its original value. A 19% decrease in flow is caused by a 5% decrease in radius. The body may compensate by increasing blood pressure by 19%, but this presents hazards to the heart and any vessel that has weakened walls. Another example comes from automobile engine oil. If you have a car with an oil pressure gauge, you may notice that oil pressure is high when the engine is cold. Motor oil has greater viscosity when cold than when warm, and so pressure must be greater to pump the same amount of cold oil.

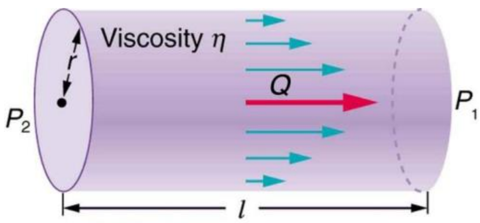


Figure 12.13: Poiseuille's law applies to laminar flow of an incompressible fluid of viscosity η through a tube of length l and radius r. The direction of flow is from greater to lower pressure. Flow rate Q is directly proportional to the pressure difference $P2-P1$, and inversely proportional to the length l of the tube and viscosity η of the fluid. Flow rate increases with $r4$, the fourth power of the radius.

Flow and Resistance as Causes of Pressure Drops

You may have noticed that water pressure in your home might be lower than normal on hot summer days when there is more use. This pressure drop occurs in the water main before it reaches your home. Let us consider flow through the water main as illustrated in Figure 12.4. We can understand why the pressure $P1$ to the home drops during times of heavy use by rearranging

$$Q = (P2-P1)/R$$

to

$$P2-P1=RQ$$

where, in this case, $P2$ is the pressure at the water works and R is the resistance of the water main. During times of heavy use, the flow rate Q is large. This means that $P2-P1$ must also be large. Thus, $P1$ must decrease. It is correct to think of flow and resistance as causing the pressure to drop from $P2$ to $P1$. $P2-P1=RQ$ is valid for both laminar and turbulent flows.

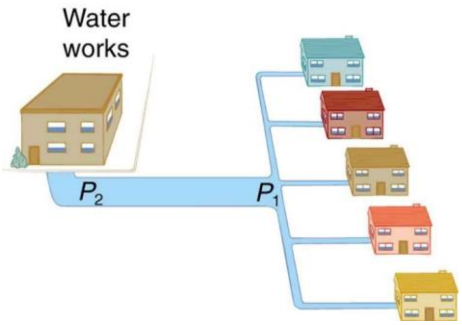


Figure 12.14: During times of heavy use, there is a significant pressure drop in a water main, and $P1$ supplied to users is significantly less than $P2$ created at the water works. If the flow is very small, then the pressure drop is negligible, and $P2 \approx P1$.

We can use $P2 - P1 = RQ$ to analyze pressure drops occurring in more complex systems in which the tube radius is not the same everywhere. Resistance will be much greater in narrow places, such as an obstructed coronary artery. For a given flow rate Q, the pressure drop will be greatest where the tube is most narrow. This is how water faucets control flow. Additionally, R is greatly increased by turbulence, and a constriction that creates turbulence greatly reduces the pressure downstream. Plaque in an artery reduces pressure and hence flow, both by its resistance and by the turbulence it creates.

Sometimes we can predict if flow will be laminar or turbulent. We know that flow in a very smooth tube or around a smooth, streamlined object will be laminar at low velocity. We also know that at high velocity, even flow in a smooth tube or around a smooth object will experience turbulence. In between, it is more difficult to predict. In fact, at intermediate velocities, flow may oscillate back and forth indefinitely between laminar and turbulent.

12.7 Reynold's Number

An occlusion, or narrowing, of an artery, such as shown in Figure 2.15, is likely to cause turbulence because of the irregularity of the blockage, as well as the complexity of blood as a fluid. Turbulence in the circulatory system is noisy and can sometimes be detected with a stethoscope, such as when measuring diastolic pressure in the upper arm's partially collapsed brachial artery. These turbulent sounds, at the onset of blood flow when the cuff pressure becomes sufficiently small, are called *Korotkoff sounds*. Aneurysms, or ballooning of arteries, create significant turbulence and can sometimes be detected with a stethoscope. Heart murmurs, consistent with their name, are sounds produced by turbulent flow around damaged and insufficiently closed heart valves. Ultrasound can also be used to detect turbulence as a medical indicator in a process analogous to Doppler-shift radar used to detect storms.

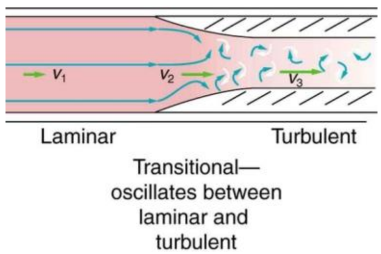


Figure 12.15: Flow is laminar in the large part of this blood vessel and turbulent in the part narrowed by plaque, where velocity is high. In the transition region, the flow can oscillate chaotically between laminar and turbulent flow.

An indicator called the Reynolds number N_R can reveal whether flow is laminar or turbulent. For flow in a tube of uniform diameter, the Reynolds number is defined as

$$N_R = (2\rho v r) / \eta \quad \text{(flow in a tube)}$$

where ρ is the fluid density, v its speed, η its viscosity, and r the tube radius. The Reynolds number is a unitless quantity. Experiments have revealed that N_R is related to the onset of turbulence. For N_R below about 2000, flow is laminar. For N_R above about 3000, flow is turbulent. For values of N_R between about 2000 and 3000, flow is unstable—that is, it can be laminar, but small obstructions and surface roughness can make it turbulent, and it may oscillate randomly between being laminar and turbulent. The blood flow through most of the body is a quiet, laminar flow. The exception is in the aorta, where the speed of the blood flow rises above a critical value of 35 m/s and becomes turbulent.

The topic of chaos has become quite popular over the last few decades. A system is defined to be *chaotic* when its behavior is so sensitive to some factor that it is extremely difficult to predict. The field of *chaos* is the study of chaotic behavior. A good example of chaotic behavior is the flow of a fluid with a Reynolds number between 2000 and 3000. Whether or not the flow is turbulent is difficult, but not impossible, to predict—the difficulty lies in the extremely sensitive dependence on factors like roughness and obstructions on the nature of the flow. A tiny variation in one factor has an exaggerated (or nonlinear) effect on the flow. Phenomena as disparate as turbulence, the orbit of Pluto, and the onset of irregular heartbeats are chaotic and can be analyzed with similar techniques.

A moving object in a viscous fluid is equivalent to a stationary object in a flowing fluid stream. (For example, when you ride a bicycle at 10 m/s in still air, you feel the air in your face exactly as if you were stationary in a 10-m/s wind.) The flow of the stationary fluid around a moving object may be laminar, turbulent, or a combination of the two. Just as with flow in tubes, it is possible to predict when a moving object creates turbulence. We use another form of the Reynolds number N'_R, defined for an object moving in a fluid to be

- $N'_R = (\rho v L) / \eta \quad$ (object in fluid) where L is a characteristic length of the object (a sphere's diameter, for example), ρ the fluid density, η its viscosity, and v the object's speed in the fluid. If N'_R is less than about 1, flow around the object can be laminar, particularly if the object has a smooth shape. The transition to turbulent flow occurs for N'_R between 1 and about 10, depending on surface roughness and so on. Depending on the surface, there can be a *turbulent wake* behind the object with some laminar flow over its surface. For an N'_R between 10 and 106, the flow may be either laminar or turbulent and may oscillate between the two. For N'_R greater than about 106, the flow is entirely turbulent, even at the surface of the object. (See Figure 12.16.) Laminar flow occurs mostly when the objects in the fluid are small, such as raindrops, pollen, and blood cells in plasma.

One of the consequences of viscosity is a resistance force called viscous drag F_V that is exerted on a moving object. This force typically depends on the object's speed (in contrast with simple friction). Experiments have shown that for laminar flow (N'_R less than about one) viscous drag is proportional to speed, whereas for N'_R between about 10 and 10^6, viscous drag is proportional to speed squared. (This relationship is a strong dependence and is pertinent to bicycle racing, where even a small headwind causes significantly increased drag on the racer. Cyclists take turns being the leader in the pack for this reason.) For N'_R greater than 10^6, drag increases dramatically and behaves with greater complexity. For laminar flow around a sphere, F_V is proportional to fluid viscosity η, the object's characteristic size L, and its speed v. All of which makes sense—the more viscous the fluid and the larger the object, the more drag we expect. (Stoke's law $F_S = 6\pi r\eta v$.) For the special case of a small sphere of radius R moving slowly in a fluid of viscosity η, the drag force F_S is given by,

$$F_S = 6\pi R\eta v$$

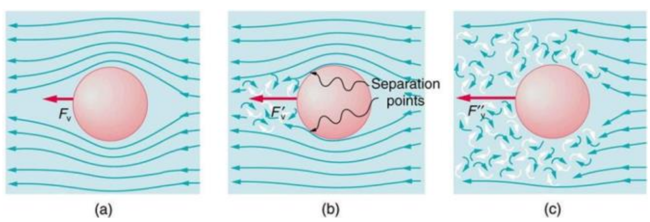


Figure 12.16: (a) Motion of this sphere to the right is equivalent to fluid flow to the left. Here the flow is laminar with N'_R less than 1. There is a force, called viscous drag F_V, to the left on the ball due to the fluid's viscosity. (b) At a higher speed, the flow becomes partially turbulent, creating a wake starting where the flow lines separate from the surface. Pressure in the wake is less than in front of the sphere, because fluid speed is less, creating a net force to the left F'_V that is significantly greater than for laminar flow. Here N'_R is greater than 10. (c) At much higher speeds, where N'_R is greater than 106, flow becomes turbulent everywhere on the surface and behind the sphere. Drag increases dramatically.

An interesting consequence of the increase in F_V with speed is that an object falling through a fluid will not continue to accelerate indefinitely (as it would if we neglect air resistance, for example). Instead, viscous drag increases, slowing acceleration, until a critical speed, called the terminal speed, is reached and the acceleration of the object becomes zero. Once this happens, the object continues to fall at constant speed (the terminal speed). This is the case for particles of sand falling in the ocean, cells falling in a centrifuge, and sky divers falling through the air. Figure 12.17 shows some of the factors that affect terminal speed. There is a viscous drag on the object that depends on the viscosity of the fluid and the size of the object. But there is also a buoyant force that depends on the density of the object relative to the fluid. Terminal speed will be greatest for low-viscosity fluids and objects with high densities and small sizes. Thus, a skydiver falls more slowly with outspread limbs than when they are in a pike position—head first with hands at their side and legs together.

Knowledge of terminal speed is useful for estimating sedimentation rates of small particles. We know from watching mud settle out of dirty water that sedimentation is usually a slow process. Centrifuges are used to speed sedimentation by creating accelerated frames in which gravitational acceleration is replaced by centripetal acceleration, which can be much greater, increasing the terminal speed.

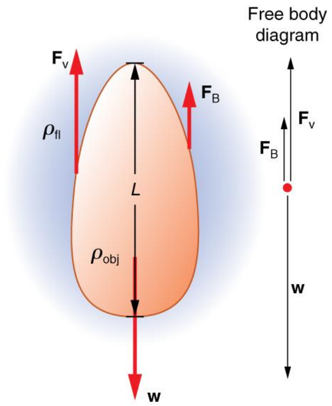


Figure 12.17: There are three forces acting on an object falling through a viscous fluid: its weight w, the viscous drag FV, and the buoyant force FB.

12.7 Diffusion

There is something fishy about the ice cube from your freezer - how did it pick up those food odors? How does soaking a sprained ankle in Epsom salt reduce swelling? The answers to these questions are related to atomic and molecular transport phenomena - another mode of fluid motion. Atoms and molecules are in constant motion at any temperature. In fluids they move about randomly even in the absence of macroscopic flow. This motion is called a random walk and is illustrated in figure 12.8. Diffusion is the movement of substances due to random thermal molecular motion. Fluids, like fish fumes or odors entering ice cubes, can even diffuse through solids.

Diffusion is a slow process over macroscopic distances. The densities of common materials are great enough that molecules cannot travel very far before having a collision that can scatter them in any direction, including straight backward. It can be shown that the average distance *x*rms that a molecule travels is proportional to the square root of time:

$$X_{\text{rms}} = 2\sqrt{Dt}$$

where x_{rms} stands for the **root-mean-square distance** and is the statistical average for the process. The quantity, *D,* is the diffusion constant for the particular molecule in a specific medium.

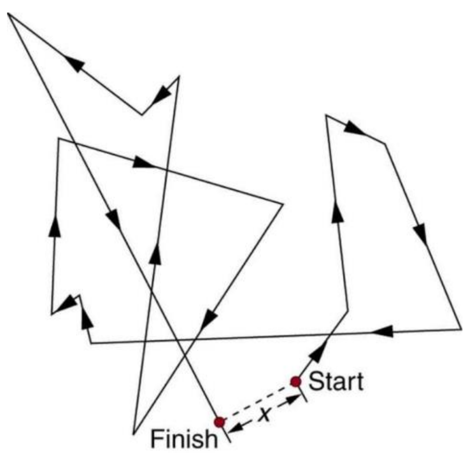


Figure 12.18: The random thermal motion of a molecule in a fluid in time t. This type of motion is called a random walk.

Note that *D* gets progressively smaller for more massive molecules. This decrease is because the average molecular speed at a given temperature is inversely proportional to molecular mass. Thus, the more massive molecules diffuse more slowly. Another interesting point is that *D* for oxygen in air is much greater than *D* for oxygen in water. In water, an oxygen molecule makes many more collisions in its random walk and is slowed considerably. In water, an oxygen molecule moves only about 40μm in 1 s. (Each molecule actually collides about 10^{10} times per second!). Finally, note that diffusion constants increase with temperature, because average molecular speed increases with temperature. This is because the average kinetic energy of molecules, $\frac{1}{2} mv^2$, is proportional to absolute temperature.

The Rate and Direction of Diffusion

If you very carefully place a drop of food coloring in a still glass of water, it will slowly diffuse into the colorless surroundings until its concentration is the same everywhere. This type of diffusion is called free diffusion, because there are no barriers inhibiting it. Let us examine its direction and rate. Molecular motion is random in direction, and so simple chance dictates that more molecules will move out of a region of high concentration than into it. The net rate of diffusion is higher initially than after the process is partially completed. (See Figure 12.19.)

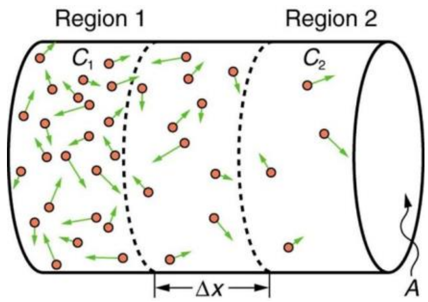


Figure 12.19: Diffusion proceeds from a region of higher concentration to a lower one. The net rate of movement is proportional to the difference in concentration.

The net rate of diffusion is proportional to the concentration difference. Many more molecules will leave a region of high concentration than will enter it from a region of low concentration. In fact, if the concentrations were the same, there would be *no* net movement. The net rate of diffusion is also proportional to the diffusion constant *D*, which is determined experimentally. The farther a molecule can diffuse in a given time, the more likely it is to leave the region of high concentration. Many of the factors that affect the rate are hidden in the diffusion constant *D*. For example, temperature and cohesive and adhesive forces all affect values of *D*.

Diffusion is the dominant mechanism by which the exchange of nutrients and waste products occur between the blood and tissue, and between air and blood in the lungs. In the evolutionary process, as organisms became larger, they needed quicker methods of transportation than net diffusion, because of the larger distances involved in the transport, leading to the development of circulatory systems. Less sophisticated, single-celled organisms still rely totally on diffusion for the removal of waste products and the uptake of nutrients.

12.8 Osmosis and Dialysis—Diffusion across Membranes

Some of the most interesting examples of diffusion occur through barriers that affect the rates of diffusion. For example, when you soak a swollen ankle in Epsom salt, water diffuses through your skin. Many substances regularly move through cell membranes; oxygen moves in, carbon dioxide moves out, nutrients go in, and wastes go out, for example. Because membranes are thin structures (typically 6.5×10^{-9} to 10×10^{-9} m across) diffusion rates through them can be high. Diffusion through membranes is an important method of transport.

Membranes are generally selectively permeable, or semipermeable. (See Figure12.20.) One type of semipermeable membrane has small pores that allow only small molecules to pass through. In other types of membranes, the molecules may actually dissolve in the membrane or react with molecules in the membrane while moving across. Membrane function, in fact, is the subject of much current research, involving not only physiology but also chemistry and physics.

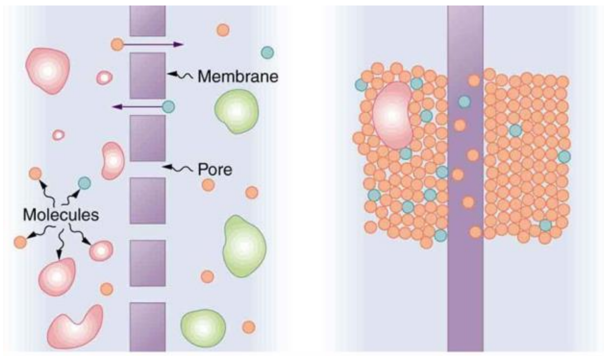


Figure 12.20: (a) A semipermeable membrane with small pores that allow only small molecules to pass through. (b) Certain molecules dissolve in this membrane and diffuse across it.

Osmosis is the transport of water through a semipermeable membrane from a region of high concentration to a region of low concentration. Osmosis is driven by the imbalance in water concentration. For example, water is more concentrated in your body than in Epsom salt. When you soak a swollen ankle in Epsom salt, the water moves out of your body into the lower-concentration region in the salt. Similarly, dialysis is the transport of any other molecule through a semipermeable membrane due to its concentration difference. Both osmosis and dialysis are used by the kidneys to cleanse the blood.

Osmosis can create a substantial pressure. Consider what happens if osmosis continues for some time, as illustrated in Figure 12.21. Water moves by osmosis from the left into the region on the right, where it is less concentrated, causing the solution on the right to rise. This movement will continue until the pressure ρgh created by the extra height of fluid on the right is large enough to stop further osmosis. This pressure is called a *back pressure*. The back pressure ρgh that stops osmosis is also called the relative osmotic pressure if neither solution is pure water, and it is called the osmotic pressure if one solution is pure water. Osmotic pressure can be large, depending on the size of the concentration difference. For example, if pure water and sea water are separated by a semipermeable membrane that passes no salt, osmotic pressure will be 25.9 atm. This value means that water will diffuse through the membrane until the salt water surface rises 268 m above the pure-water surface! One example of pressure created by osmosis is turgor in plants (many wilt when too dry). Turgor describes the condition of a plant in which the fluid in a cell exerts a pressure against the cell wall. This pressure gives the plant support. Dialysis can similarly cause substantial pressures.

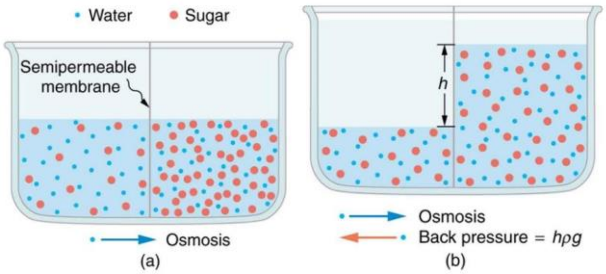


Figure 12.21: (a) Two sugar-water solutions of different concentrations, separated by a semipermeable membrane that passes water but not sugar. Osmosis will be to the right, since water is less concentrated there. (b) The fluid level rises until the back pressure ρgh equals the relative osmotic pressure; then, the net transfer of water is zero.

Reverse osmosis and reverse dialysis (also called filtration) are processes that occur when back pressure is sufficient to reverse the normal direction of substances through membranes. Back pressure can be created naturally as on the right side of Figure 12.21. (A piston can also create this pressure.) Reverse osmosis can be used to desalinate water by simply forcing it through a membrane that will not pass salt. Similarly, reverse dialysis can be used to filter out any substance that a given membrane will not pass.

One further example of the movement of substances through membranes deserves mention. We sometimes find that substances pass in the direction opposite to what we expect. Cypress tree roots, for example, extract pure water from salt water, although osmosis would move it in the opposite direction. This is not reverse osmosis, because there is no back pressure to cause it. What is happening is called active transport, a process in which a living membrane expends energy to move substances across it. Many living membranes move water and other substances by active transport. The kidneys, for example, not only use osmosis and dialysis—they also employ significant active transport to move substances into and out of blood. In fact, it is estimated that at least 25% of the body's energy is expended on active transport of substances at the cellular level. The study of active transport carries us into the realms of microbiology, biophysics, and biochemistry and it is a fascinating application of the laws of nature to living structures.

For examples and answers, please refer to OpenStax.com questions and answers given on their website or to the College Physics 2e - https://openstax.org/details/books/college-physics-2e.

Chapter 13

13.0 Objectives

At the end of this lesson, students should be able to,

- Identify boiling and evaporation.
- Convert temperatures to different scales.
- Calculate thermal expansion.
- Apply ideal gas laws and kinetic theory to solve problems related to gases.
- Identify phase changes.
- Calculate relative humidity.

13.1 Introduction

This chapter covers boiling, evaporation, temperature conversions, thermal expansion, ideal gas laws, kinetic theory, phase changes, and relative humidity and related calculations.

13.2 Temperature

The concept of temperature has evolved from the common concepts of hot and cold. Human perception of what feels hot or cold is a relative one. For example, if you place one hand in hot water and the other in cold water, and then place both hands in tepid water, the tepid water will feel cool to the hand that was in hot water, and warm to the one that was in cold water. The scientific definition of temperature is less ambiguous than your senses of hot and cold. Temperature is operationally defined to be what we measure with a thermometer. (Many physical quantities are defined solely in terms of how they are measured. We shall see later how temperature is related to the kinetic energies of atoms and molecules, a more physical explanation.) Two accurate thermometers, one placed in hot water and the other in cold water, will show the hot water to have a higher temperature. If they are then placed in tepid water, both will give identical readings (within measurement uncertainties). In this section, we discuss temperature, its measurement by thermometers, and its relationship to thermal equilibrium. Again, temperature is the quantity measured by a thermometer.

Any physical property that depends on temperature, and whose response to temperature is reproducible, can be used as the basis of a thermometer. Because many physical properties depend on temperature, the variety of thermometers is remarkable. For example, volume increases with temperature for most substances. This property is the basis for the common alcohol thermometer, the old mercury thermometer, and the bimetallic strip (Figure 13.1). Other properties used to measure temperature include electrical resistance and color, as shown in Figure 13.2, and the emission of infrared radiation, as shown in Figure 13.3.

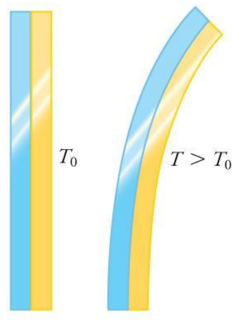


Figure 13.1: The curvature of a bimetallic strip depends on temperature. (a) The strip is straight at the starting temperature, where its two components have the same length. (b) At a higher temperature, this strip bends to the right, because the metal on the left has expanded more than the metal on the right.

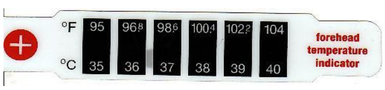


Figure 13.2: Each of the six squares on this plastic (liquid crystal) thermometer contains a film of a different heat-sensitive liquid crystal material. Below 95ºF, all six squares are black. When the plastic thermometer is exposed to temperature that increases to 95ºF, the first liquid crystal square changes color. When the temperature increases above 96.8ºF the second liquid crystal square also changes color, and so forth. (credit: Arkrishna, Wikimedia Commons)

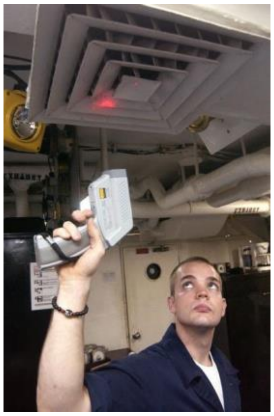


Figure 13.3: Fireman Jason Ormand uses a pyrometer to check the temperature of an aircraft carrier's ventilation system. Infrared radiation (whose emission varies with temperature) from the vent is measured and a temperature readout is quickly produced. Infrared measurements are also frequently used as a measure of body temperature. These modern thermometers, placed in

the ear canal, are more accurate than alcohol thermometers placed under the tongue or in the armpit. (credit: Lamel J. Hinton/U.S. Navy)

13.3 Temperature Scales

Thermometers are used to measure temperature according to well-defined scales of measurement, which use pre-defined reference points to help compare quantities. The three most common temperature scales are the Fahrenheit, Celsius, and Kelvin scales. A temperature scale can be created by identifying two easily reproducible temperatures. The freezing and boiling temperatures of water at standard atmospheric pressure are commonly used.

The Celsius scale (which replaced the slightly different *centigrade* scale) has the freezing point of water at 0ºC and the boiling point at 100ºC. Its unit is the degree Celsius(ºC). On the Fahrenheit scale (still the most frequently used in the United States), the freezing point of water is at 32ºF and the boiling point is at 212ºF. The unit of temperature on this scale is the degree Fahrenheit(ºF). Note that a temperature difference of one degree Celsius is greater than a temperature difference of one degree Fahrenheit. Only 100 Celsius degrees span the same range as 180 Fahrenheit degrees, thus one degree on the Celsius scale is 1.8 times larger than one degree on the Fahrenheit scale 180/100 = 9/5.

The Kelvin scale is the temperature scale that is commonly used in science. It is an *absolute temperature* scale defined to have 0 K at the lowest possible temperature, called absolute zero. The official temperature unit on this scale is the *kelvin*, which is abbreviated K, and is not accompanied by a degree sign. The freezing and boiling points of water are 273.15 K and 373.15 K, respectively. Thus, the magnitude of temperature differences is the same in units of kelvins and degrees Celsius. Unlike other temperature scales, the Kelvin scale is an absolute scale. It is used extensively in scientific work because a number of physical quantities, such as the volume of an ideal gas, are directly related to absolute temperature. The kelvin is the SI unit used in scientific work.

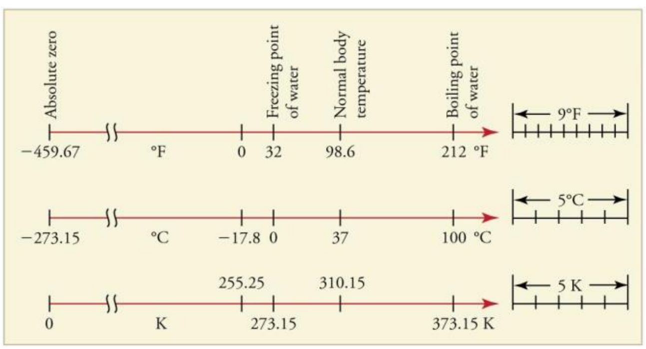


Figure 13.4: Relationships between the Fahrenheit, Celsius, and Kelvin temperature scales, rounded to the nearest degree. The relative sizes of the scales are also shown.

The relationships between the three common temperature scales is shown in Figure 13.4.. Temperatures on these scales can be converted using the equations in Table 13.1.

To convert from . . .	Use this equation . . .	Also written as . . .
Celsius to Fahrenheit	$T(^\circ F)=(9/5)T(^\circ C)+32$	$T^\circ F=(9/5)T^\circ C+32$
Fahrenheit to Celsius	$T(^\circ C)=(5/9)(T(^\circ F)-32)$	$T^\circ C=(5/9)(T^\circ F-32)$
Celsius to Kelvin	$T(K)=T(^\circ C)+273.15$	$TK=T^\circ C+273.15$
Kelvin to Celsius	$T(^\circ C)=T(K)-273.15$	$T^\circ C=TK-273.15$
Fahrenheit to Kelvin	$T(K)=(5/9)(T(^\circ F)-32)+273.15$	$TK=(5/9)(T^\circ F-32)+273.15$
Kelvin to Fahrenheit	$T(^\circ F)=(9/5)(T(K)-273.15)+32$	$T^\circ F=(9/5)(TK-273.15)+32$

Table 13.1: Temperature Conversions

Notice that the conversions between Fahrenheit and Kelvin look quite complicated. In fact, they are simple combinations of the conversions between Fahrenheit and Celsius, and the conversions between Celsius and Kelvin.

Example - Converting between Temperature Scales: Room Temperature

"Room temperature" is generally defined to be 25ºC. (a) What is the room temperature in ºF? (b) What is it in K?

Strategy

To answer these questions, all we need to do is choose the correct conversion equations and plug in the known values.

Solution for (a)

1. Choose the right equation. To convert from ºC to ºF, use the equation,

$$T^\circ F = (9/5)T^\circ C+32.$$

2. Plug the known value into the equation and solve:

$$T^\circ F = (9/5)25^\circ C+32 = 77^\circ F$$

Solution for (b)

1. Choose the right equation. To convert from ºC to K, use the equation,

$$T\text{K}=T\text{ºC}+273.15.$$

2. Plug the known value into the equation and solve:

$$T\text{K}=25\text{ºC}+273.15=298\text{K}.$$

Example - Converting between Temperature Scales: the Reaumur Scale

The Reaumur scale is a temperature scale that was used widely in Europe in the 18th and 19th centuries. On the Reaumur temperature scale, the freezing point of water is 0ºR and the boiling temperature is 80ºR. If "room temperature" is 25ºC on the Celsius scale, what is it on the Reaumur scale?

Strategy

To answer this question, we must compare the Reaumur scale to the Celsius scale. The difference between the freezing point and boiling point of water on the Reaumur scale is 80ºR. On the Celsius scale it is 100ºC. Therefore 100ºC=80ºR. Both scales start at 0º for freezing, so we can derive a simple formula to convert between temperatures on the two scales.

Solution

1. Derive a formula to convert from one scale to the other:

$$T\text{ºR} = (0.8\text{ºR}^{o} \,/\, \text{C}) \times T\text{ºC}.$$

2. Plug the known value into the equation and solve:

$$T\text{ºR} = (0.8\text{ºR} /\text{ºC}) \times 25\text{ºC} = 20\text{ºR}$$

Absolute Zero

What is absolute zero? Absolute zero is the temperature at which all molecular motion has ceased. The concept of absolute zero arises from the behavior of gases. Figure 13.5 shows how the pressure of gases at a constant volume decrease as temperature decreases. Various scientists have noted that the pressures of gases extrapolate to zero at the same temperature, –273.15ºC. This extrapolation implies that there is a lowest temperature. This temperature is called *absolute zero*. Today we know that most gases first liquefy and then freeze, and it is not actually possible to reach absolute zero. The numerical value of absolute zero temperature is –273.15ºC or 0 K.

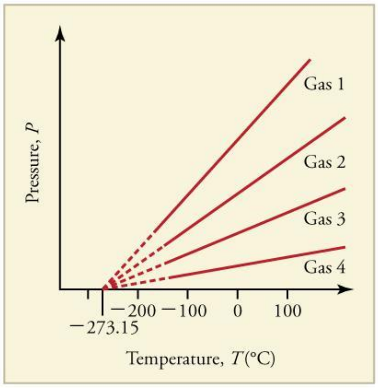


Figure 13.5: Graph of pressure versus temperature for various gases kept at a constant volume. Note that all of the graphs extrapolate to zero pressure at the same temperature.

13.4 Thermal Equilibrium and the Zeroth Law of Thermodynamics

Thermometers actually take their *own* temperature, not the temperature of the object they are measuring. This raises the question of how we can be certain that a thermometer measures the temperature of the object with which it is in contact. It is based on the fact that any two systems placed in *thermal contact* (meaning heat transfer can occur between them) will reach the same temperature. That is, heat will flow from the hotter object to the cooler one until they have exactly the same temperature. The objects are then in thermal equilibrium, and no further changes will occur. The systems interact and change because their temperatures differ, and the changes stop once their temperatures are the same. Thus, if enough time is allowed for this transfer of heat to run its course, the temperature a thermometer registers *does* represent the system with which it is in thermal equilibrium. Thermal equilibrium is established when two bodies are in contact with each other and can freely exchange energy.

Furthermore, experimentation has shown that if two systems, A and B, are in thermal equilibrium with each another, and B is in thermal equilibrium with a third system C, then A is also in thermal equilibrium with C. This conclusion may seem obvious, because all three have the same temperature, but it is basic to thermodynamics. It is called the zeroth law of thermodynamics.

This law was postulated in the 1930s, after the first and second laws of thermodynamics had been developed and named. It is called the *zeroth law* because it comes logically before the first and second laws. Suppose, for example, a cold metal block and a hot metal block are both placed on a metal plate at room temperature. Eventually the cold block and the plate will be in thermal equilibrium. In addition, the hot block and the plate will be in thermal equilibrium. By the zeroth law, we can conclude that the cold block and the hot block are also in thermal equilibrium.

The expansion of alcohol in a thermometer is one of many commonly encountered examples of thermal expansion, the change in size or volume of a given mass with temperature. Hot air rises because its volume increases, which causes the hot air’s density to be smaller than the density of surrounding air, causing a buoyant (upward) force on the hot air. The same happens in all liquids and gases, driving natural heat transfer upwards in homes, oceans, and weather systems. Solids also undergo thermal expansion. Railroad tracks and bridges, for example, have expansion joints to allow them to freely expand and contract with temperature changes.

What are the basic properties of thermal expansion? First, thermal expansion is clearly related to temperature change. The greater the temperature change, the more a bimetallic strip will bend. Second, it depends on the material. In a thermometer, for example, the expansion of alcohol is much greater than the expansion of the glass containing it.

What is the underlying cause of thermal expansion? An increase in temperature implies an increase in the kinetic energy of the individual atoms. In a solid, unlike in a gas, the atoms or molecules are closely packed together, but their kinetic energy (in the form of small, rapid vibrations) pushes neighboring atoms or molecules apart from each other. This neighbor-to-neighbor pushing results in a slightly greater distance, on average, between neighbors, and adds up to a larger size for the whole body. For most substances under ordinary conditions, there is no preferred direction, and an increase in temperature will increase the solid's size by a certain fraction in each dimension.

13.5 Linear Thermal Expansion—Thermal Expansion in One Dimension

The change in length ΔL is proportional to length L. The dependence of thermal expansion on temperature, substance, and length is summarized in the equation,

$$\Delta L = \alpha L \Delta T$$

where ΔL is the change in length L, ΔT is the change in temperature, and α is the coefficient of linear expansion, which varies slightly with temperature.

Table 13.2 lists representative values of the coefficient of linear expansion, which may have units of 1/ºC or 1/K. Because the size of a kelvin and a degree Celsius are the same, both α and ΔT can be expressed in units of kelvins or degrees Celsius. The equation $\Delta L = \alpha L \Delta T$ is accurate for small changes in temperature and can be used for large changes in temperature if an average value of α is used.

Material	**Coefficient of linear expansion α(1/ºC)**	**Coefficient of volume expansion β(1/ºC)**
Solids		
Aluminum	25×10^{-6}	75×10^{-6}
Brass	19×10^{-6}	56×10^{-6}
Copper	17×10^{-6}	51×10^{-6}
Gold	14×10^{-6}	42×10^{-6}
Iron or Steel	12×10^{-6}	35×10^{-6}
Invar (Nickel-iron alloy)	0.9×10^{-6}	2.7×10^{-6}

Material	Coefficient of linear expansion α(1/ºC)	Coefficient of volume expansion β(1/ºC)
Lead	29×10^{-6}	87×10^{-6}
Silver	18×10^{-6}	54×10^{-6}
Glass (ordinary)	9×10^{-6}	27×10^{-6}
Glass (Pyrex®)	3×10^{-6}	9×10^{-6}
Quartz	0.4×10^{-6}	1×10^{-6}
Concrete, Brick	$\sim12\times10^{-6}$	$\sim36\times10^{-6}$
Marble (average)	7×10^{-6}	2.1×10^{-5}
Liquids		
Ether		1650×10^{-6}
Ethyl alcohol		1100×10^{-6}
Petrol		950×10^{-6}
Glycerin		500×10^{-6}
Mercury		180×10^{-6}
Water		210×10^{-6}
Gases		
Air and most other gases at atmospheric pressure		3400×10^{-6}

Table 13.2: Thermal Expansion Coefficients at 20ºC

Example - Calculating Linear Thermal Expansion: The Golden Gate Bridge

The main span of San Francisco's Golden Gate Bridge is 1275 m long at its coldest. The bridge is exposed to temperatures ranging from –15ºC to 40ºC. What is its change in length between these temperatures? Assume that the bridge is made entirely of steel.

Strategy

Use the equation for linear thermal expansion $\Delta L=\alpha L\Delta T$ to calculate the change in length , ΔL. Use the coefficient of linear expansion, α, for steel from Table 13.2, and note that the change in temperature, ΔT, is 55ºC.

Solution

Plug all of the known values into the equation to solve for ΔL.

$$\Delta L=\alpha L\Delta T=(12\times10^{-6}\ /\ ºC)(1275\ m)(55ºC) = 0.84\ m.$$

Discussion

Although not large compared with the length of the bridge, this change in length is observable. It is generally spread over many expansion joints so that the expansion at each joint is small.

13.6 Thermal Expansion in Two and Three Dimensions

Objects expand in all dimensions, as illustrated in Figure 13.6. That is, their areas and volumes, as well as their lengths, increase with temperature. Holes also get larger with temperature. If you cut a hole in a metal plate, the remaining material will expand exactly as it would if the plug was still in place. The plug would get bigger, and so the hole must get bigger too. (Think of the ring of neighboring atoms or molecules on the wall of the hole as pushing each other farther apart as temperature increases. Obviously, the ring of neighbors must get slightly larger, so the hole gets slightly larger).

13.6.1 Thermal Expansion in Two Dimensions

For small temperature changes, the change in area ΔA is given by,

$$\Delta A = 2\alpha A \Delta T,$$

where ΔA is the change in area A, ΔT is the change in temperature, and α is the coefficient of linear expansion, which varies slightly with temperature.

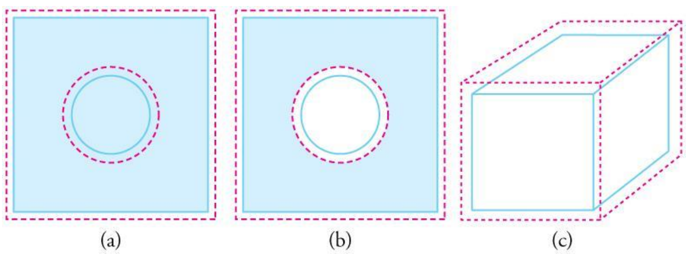


Figure 13.6: In general, objects expand in all directions as temperature increases. In these drawings, the original boundaries of the objects are shown with solid lines, and the expanded boundaries with dashed lines. (a) Area increases because both length and width increase. The area of a circular plug also increases. (b) If the plug is removed, the hole it leaves becomes larger with increasing temperature, just as if the expanding plug were still in place. (c) Volume also increases, because all three dimensions increase.

13.6.2 Thermal Expansion in Three Dimensions

The change in volume ΔV is very nearly $\Delta V = 3\alpha V \Delta T$. This equation is usually written as

$$\Delta V = \beta V \Delta T,$$

where β is the coefficient of volume expansion and $\beta \approx 3\alpha$. Note that the values of β in Table 13.2 are almost exactly equal to 3α.

In general, objects will expand with increasing temperature. Water is the most important exception to this rule. Water expands with increasing temperature (its density *decreases*) when it is at temperatures greater than 4ºC(40ºF). However, it expands with *decreasing* temperature when it is between +4ºC and 0ºC (40ºF to 32ºF). Water is densest at +4ºC. (See Figure 13.7) Perhaps the most striking effect of this phenomenon is the freezing of water in a pond. When water near the surface cools down to 4ºC it is denser than the remaining water and thus will sink to the bottom. This "turnover" results in a layer of warmer water near the surface, which is then cooled. Eventually the pond has a uniform temperature of 4ºC. If the temperature in the surface layer drops below 4ºC, the water is less dense than the water below, and thus stays near the top. As a result, the pond surface can completely freeze over. The ice on top of liquid water provides an insulating layer from winter's harsh exterior air temperatures. Fish and other aquatic life can survive in 4ºC water beneath ice, due to this unusual characteristic of water. It also produces circulation of water in the pond that is necessary for a healthy ecosystem of the body of water.

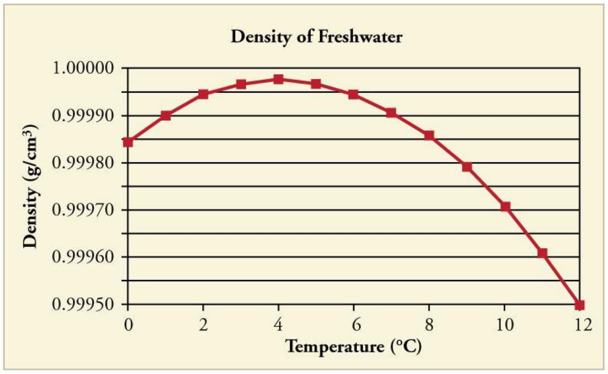


Figure 13.7: The density of water as a function of temperature. Note that the thermal expansion is actually very small. The maximum density at +4ºC is only *0.0075%* greater than the density at 2ºC, and *0.012%* greater than that at 0ºC.

Example - Calculating Thermal Expansion: Gas vs. Gas Tank

Suppose your 60.0-L (15.9-gal) steel gasoline tank is full of gas, so both the tank and the gasoline have a temperature of 15.0ºC. How much gasoline has spilled by the time they warm to 35.0ºC?

Strategy

The tank and gasoline increase in volume, but the gasoline increases more, so the amount spilled is the difference in their volume changes. (The gasoline tank can be treated as solid steel.) We can use the equation for volume expansion to calculate the change in volume of the gasoline and of the tank.

Solution

1. Use the equation for volume expansion to calculate the increase in volume of the steel tank:

$$\Delta V_s = \beta_s V_s \Delta T$$

2. The increase in volume of the gasoline is given by this equation:

$$\Delta V_{gas} = \beta_{gas} V_{gas} \Delta T$$

3. Find the difference in volume to determine the amount spilled as

$$V_{spill} = \Delta V_{gas} - \Delta V_s$$

Alternatively, we can combine these three equations into a single equation. (Note that the original volumes are equal.)

$$V_{spill} = (\beta_{gas} - \beta_s) V \Delta T$$

$$= [(950 - 35) \times 10^{-6}/^{\circ}C](60.0 L)(20.0^{\circ}C)$$

$$= 1.10 L$$

Discussion

This amount is significant, particularly for a 60.0-L tank. The effect is so striking because gasoline and steel expand quickly.

If you try to cap the tank tightly to prevent overflow, you will find that it leaks anyway, either around the cap or by bursting the tank. Tightly constricting the expanding gas is equivalent to compressing it, and both liquids and solids resist being compressed with extremely large forces. To avoid rupturing rigid containers, these containers have air gaps, which allow them to expand and contract without stressing them.

13.7 Thermal Stress

Thermal stress is created by thermal expansion or contraction. Thermal stress can be destructive, such as when expanding gasoline ruptures a tank. It can also be useful, for example, when two parts are joined together by heating one in manufacturing, then slipping it over the other and allowing the combination to cool. Thermal stress can explain many phenomena, such as the weathering of rocks and pavement by the expansion of ice when it freezes.

Example - Calculating Thermal Stress: Gas Pressure

What pressure would be created in the gasoline tank considered in the previous example, if the gasoline increases in temperature from 15.0ºC to 35.0ºC without being allowed to expand? Assume that the bulk modulus B for gasoline is $1.00\times10^{9} N/m^{2}$.

Strategy

To solve this problem, we must use the following equation, which relates a change in volume ΔV to pressure:

$$\Delta V = (1/B)\ (F/A)\ V_0$$

where F/A is pressure, V_0 is the original volume, and B is the bulk modulus of the material involved. We will use the amount spilled in previous example as the change in volume, ΔV.

Solution

1. Rearrange the equation for calculating pressure:

$$P = F/A = (\Delta V/V_0)\ B$$

2. Insert the known values. The bulk modulus for gasoline is $B = 1.00\times10^9 \text{N/m}^2$. In the previous example, the change in volume $\Delta V = 1.10\text{L}$ is the amount that would spill. Here, $V_0 = 60.0\text{L}$ is the original volume of the gasoline. Substituting these values into the equation, we obtain

$$P = (1.10\ \text{L} / 60.0\ \text{L})\ (1.00\times10^9 \text{Pa}) = 1.83\times10^7 \text{Pa}.$$

Discussion

This pressure is about 2500lb/in^2, *much* more than a gasoline tank can handle.

Forces and pressures created by thermal stress are typically as great as that in the example above. Railroad tracks and roadways can buckle on hot days if they lack sufficient expansion joints. (See Figure 13.8). Power lines sag more in the summer than in the winter, and will snap in cold weather if there is insufficient slack. Cracks open and close in plaster walls as a house warms and cools. Glass cooking pans will crack if cooled rapidly or unevenly, because of differential contraction and the stresses it creates. (Pyrex® is less susceptible because of its small coefficient of thermal expansion.) Nuclear reactor pressure vessels are threatened by overly rapid cooling, and although none have failed, several have been cooled faster than considered desirable. Biological cells are ruptured when foods are frozen, detracting from their taste. Repeated thawing and freezing accentuate the damage. Even the oceans can be affected. A significant portion of the rise in sea level that is resulting from global warming is due to the thermal expansion of sea water.

Figure 13.8: Thermal stress contributes to the formation of potholes. (credit: Editor5807, Wikimedia Commons)

Metal is regularly used in the human body for hip and knee implants. Most implants need to be replaced over time because, among other things, metal does not bond with bone. Researchers are trying to find better metal coatings that would allow metal-to-bone bonding. One challenge is to find a coating that has an expansion coefficient similar to that of metal. If the expansion coefficients are too different, the thermal stresses during the manufacturing process leads to cracks at the coating-metal interface.

Another example of thermal stress is found in the mouth. Dental fillings can expand differently from tooth enamel. It can give pain when eating ice cream or having a hot drink. Cracks might occur in the filling. Metal fillings (gold, silver, etc.) are being replaced by composite fillings (porcelain), which have smaller coefficients of expansion, and are closer to those of teeth.

Gases are easily compressed. We can see evidence of this in Table 13.2, where you will note that gases have the *largest* coefficients of volume expansion. The large coefficients mean that gases expand and contract very rapidly with temperature changes. In addition, you will note that most gases expand at the *same* rate, or have the same β. This raises the question as to why gases should all act in nearly the same way, when liquids and solids have widely varying expansion rates.

The answer lies in the large separation of atoms and molecules in gases, compared to their sizes, as illustrated in Figure13.9. Because atoms and molecules have large separations, forces between them can be ignored, except when they collide with each other during collisions. The motion of atoms and molecules (at temperatures well above the boiling temperature) is fast, such that the gas occupies all of the accessible volume and the expansion of gases is rapid. In contrast, in liquids and solids, atoms and molecules are closer together and are quite sensitive to the forces between them.

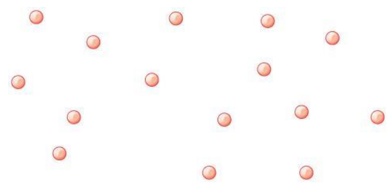


Figure 13.9: Atoms and molecules in a gas are typically widely separated, as shown. Because the forces between them are quite weak at these distances, the properties of a gas depend more on the number of atoms per unit volume and on temperature than on the type of atom.

To get some idea of how pressure, temperature, and volume of a gas are related to one another, consider what happens when you pump air into an initially deflated tire. The tire's volume first increases in direct proportion to the amount of air injected, without much increase in the tire pressure. Once the tire has expanded to nearly its full size, the walls limit volume expansion. If we continue to pump air into it, the pressure increases. The pressure will further increase when the car is driven and the tires move. Most manufacturers specify optimal tire pressure for cold tires. (See Figure 13.10.)

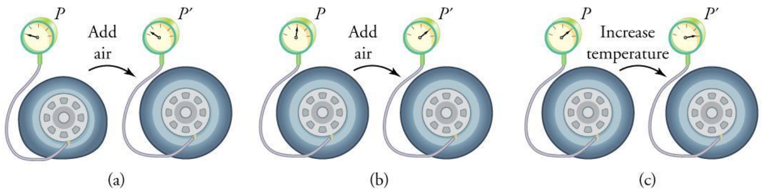


Figure 13.10: (a) When air is pumped into a deflated tire, its volume first increases without much increase in pressure. (b) When the tire is filled to a certain point, the tire walls resist further expansion, and the pressure increases with more air. (c) Once the tire is inflated, its pressure increases with temperature.

At room temperatures, collisions between atoms and molecules can be ignored. In this case, the gas is called an ideal gas, in which case the relationship between the pressure, volume, and temperature is given by the equation of state called the ideal gas law.

13.8 Ideal Gas Law

The ideal gas law states that

$$PV = NkT$$

where P is the absolute pressure of a gas, V is the volume it occupies, N is the number of atoms and molecules in the gas, and T is its absolute temperature. The constant k is called the Boltzmann constant in honor of Austrian physicist Ludwig Boltzmann (1844–1906) and has the value

$$k = 1.38 \times 10^{-23} \text{J/K}$$

The ideal gas law can be derived from basic principles, but was originally deduced from experimental measurements of Charles' law (that volume occupied by a gas is proportional to temperature at a fixed pressure) and from Boyle's law (that for a fixed temperature, the product PV is a constant). In the ideal gas model, the volume occupied by its atoms and molecules is a negligible fraction of V. The ideal gas law describes the behavior of real gases under most conditions. (Note, for example, that N is the total number of atoms and molecules, independent of the type of gas.)

Let us see how the ideal gas law is consistent with the behavior of filling the tire when it is pumped slowly and the temperature is constant. At first, the pressure P is essentially equal to atmospheric pressure, and the volume V increases in direct proportion to the number of atoms and molecules N put into the tire. Once the volume of the tire is constant, the equation $PV=NkT$ predicts that the pressure should increase in proportion to *the number N of atoms and molecules*.

Example - Calculating Pressure Changes Due to Temperature Changes: Tire Pressure

Suppose your bicycle tire is fully inflated, with an absolute pressure of 7.00×10^5Pa (a gauge pressure of just under 90.0lb/in^2) at a temperature of 18.0ºC. What is the pressure after its temperature has risen to 35.0ºC? Assume that there are no appreciable leaks or changes in volume.

Strategy

The pressure in the tire is changing only because of changes in temperature. Firstly, we need to identify what we know and what we want to know, and then identify an equation to solve for the unknown.

We know the initial pressure $P_0 = 7.00\times10^5$ Pa, the initial temperature $T_0 = 18.0$ºC, and the final temperature $T_f = 35.0$ºC. We must find the final pressure P_f. How can we use the equation $PV = NkT$? At first, it may seem that not enough information is given, because the volume V and number of atoms N are not specified. What we can do is use the equation twice: $P_0V_0 = NkT_0$ and $P_fV_f = NkT_f$. If we divide P_fV_f by P_0V_0 we can come up with an equation that allows us to solve for P_f.

$$(P_fV_f) / (P_0V_0) = (N_fkT_f) / (N_0kT_0)$$

Since the volume is constant, V_f and V_0 are the same and they cancel out. The same is true for N_f and N_0, and k, which is a constant. Therefore,

$$P_f / P_0 = T_f / T_0$$

We can then rearrange this to solve for P_f:

$$P_f = (P_0T_f) / T_0$$

where the temperature must be in units of kelvins, because T_0 and T_f are absolute temperatures.

Solution

1. Convert temperatures from Celsius to Kelvin.

$$T_0 = (18.0+273)\text{K} = 291\ \text{K}$$
$$T_f = (35.0+273)\text{K} = 308\ \text{K}$$

2. Substitute the known values into the equation.

$$P_f = P_0(T_f\ /\ T_0) = 7.00\times10^5\text{Pa}\ (308\ \text{K}\ /\ 291\ \text{K}) = 7.41\times10^5\text{Pa}$$

Discussion

The final temperature is about 6% greater than the original temperature, so the final pressure is about 6% greater as well. Note that *absolute* pressure and *absolute* temperature must be used in the ideal gas law.

Example - Calculating the Number of Molecules in a Cubic Meter of Gas

How many molecules are in a typical object, such as gas in a tire or water in a drink? We can use the ideal gas law to give us an idea of how large *N* typically is.

Calculate the number of molecules in a cubic meter of gas at standard temperature and pressure (STP), which is defined to be 0ºC and atmospheric pressure.

Strategy

Because pressure, volume, and temperature are all specified, we can use the ideal gas law $PV=NkT$, to find *N*.

Solution

1. Identify the knowns.

$$T = 0\text{ºC} = 273\ \text{K}$$
$$\text{P} = 1.01\times10^5\text{Pa}$$
$$\text{V} = 1.00\ \text{m}^3$$
$$\text{K} = 1.38\times10^{-23}\ \text{J/K}$$

2. Identify the unknown: number of molecules, *N*.

3. Rearrange the ideal gas law to solve for *N*.

$$PV = NkT$$

$$N = (PV)\ /\ (kT)$$

4. Substitute the known values into the equation and solve for N.

$$N = (PV) / (kT) = [(1.01\times10^{5}\text{Pa})(1.00\ \text{m}^{3})] / [(1.38\times10^{-23}\text{J/K})(273\ \text{K})\ ] = 2.68\times10^{25}\text{molecules}$$

Discussion

This number is undeniably large, considering that a gas is mostly empty space. N is huge, even in small volumes. For example, 1cm^3 of a gas at STP has 2.68×10^{19} molecules in it. Once again, note that N is the same for all types or mixtures of gases.

13.9 Moles and Avogadro's Number

It is sometimes convenient to work with a unit other than molecules when measuring the amount of substance. A mole (abbreviated mol) is defined to be the amount of a substance that contains as many atoms or molecules as there are atoms in exactly 12 grams (0.012 kg) of carbon-12. The actual number of atoms or molecules in one mole is called Avogadro's number(NA), in recognition of Italian scientist Amedeo Avogadro (1776–1856). He developed the concept of the mole, based on the hypothesis that equal volumes of gas, at the same pressure and temperature, contain equal numbers of molecules. That is, the number is independent of the type of gas. This hypothesis has been confirmed, and the value of Avogadro's number is

$$N_A = 6.02\times10^{23}\text{mol}^{-1}$$

Avogadro's Number

One mole always contains 6.02×10^{23} particles (atoms or molecules), independent of the element or substance. A mole of any substance has a mass in grams equal to its molecular mass, which can be calculated from the atomic masses given in the periodic table of elements.

$$N_A = 6.02\times10^{23}\text{mol}^{-1}$$

Example - Calculating Moles per Cubic Meter and Liters per Mole

Calculate: (a) the number of moles in 1.00m^3 of gas at STP, and (b) the number of liters of gas per mole.

Strategy and Solution

(a) We are asked to find the number of moles per cubic meter, and we know from the previous example that the number of molecules per cubic meter at STP is 2.68×10^{25}. The number of moles can be found by dividing the number of molecules by Avogadro's number. We let n stand for the number of moles,

$$n\ \text{mol} / \text{m}^3 = (N\ \text{molecules} / \text{m}^3) / (6.02\times10^{23}\text{moleculesc/mol})$$

$$= (2.68\times10^{25}\text{molecules/m}^3) / (6.02\times10^{23}\text{molecules/mol})$$

$$= 44.5\text{mol/m}^3$$

(b) Using the value obtained for the number of moles in a cubic meter, and converting cubic meters to liters, we obtain

$$(10^3\text{L/m}^3) / (44.5\text{mol/m}^3) = 22.5 \text{ L/mol}$$

Discussion

This value is very close to the accepted value of 22.4 L/mol. The slight difference is due to rounding errors caused by using three-digit input. Again, this number is the same for all gases. In other words, it is independent of the gas.

The (average) molar weight of air (approximately) 80% N_2 and 20% O_2 is M = 28.8g. Thus, the mass of one cubic meter of air is 1.28 kg. If a living room has dimensions 5m×5 m×3m, the mass of air inside the room is 96 kg, which is the typical mass of a human.

The Ideal Gas Law Restated Using Moles

A very common expression of the ideal gas law uses the number of moles, n, rather than the number of atoms and molecules, N. We start from the ideal gas law,

$$PV = NkT$$

and multiply and divide the equation by Avogadro's number N_A. This gives,

$$PV = (N/N_A)\, N_A kT$$

Note that $n = (N/N_A)$ is the number of moles. We define the universal gas constant $R = N_A k$, and obtain the ideal gas law in terms of moles.

Ideal Gas Law (in terms of moles)

The ideal gas law (in terms of moles) is,

$$PV = nRT$$

The numerical value of R in SI units is,

$$R = N_A k = (6.02\times10^{23}\text{mol}^{-1})(1.38\times10^{-23}\text{J/K}) = 8.31\text{J/mol}\cdot\text{K}.$$

In other units,

$$R = 1.99 \text{ cal/mol}\cdot\text{K}$$

$$R = 0.0821\ \text{L·atm/mol·K}$$

You can use whichever value of *R* is most convenient for a particular problem.

Example - Calculating Number of Moles: Gas in a Bike Tire

How many moles of gas are in a bike tire with a volume of $2.00 \times 10^{-3} m^3$(2.00 L), a pressure of 7.00×10^5Pa (a gauge pressure of just under 90.0lb/in^2), and at a temperature of 18.0ºC?

Strategy

Identify the knowns and unknowns and choose an equation to solve for the unknown. In this case, we solve the ideal gas law, $PV = nRT$, for the number of moles *n*.

Solution

1. Identify the knowns.

$$P = 7.00 \times 10^5 \text{Pa}$$
$$V = 2.00 \times 10^{-3} \text{m}^3$$
$$T = 18.0\ ºC = 291\ K$$
$$R = 8.31\ \text{J/mol·K}$$

2. Rearrange the equation to solve for *n* and substitute known values.

$$n = (PV) / (RT) = [(7.00 \times 10^5 \text{Pa})\ (2.00 \times 10^{-3} \text{m}^3)] / [(8.31 \text{J/mol·K})(291 \text{K})\ ]$$

$$= 0.579 \text{mol}$$

Discussion

The most convenient choice for *R* in this case is 8.31J/mol·K, because our known quantities are in SI units.

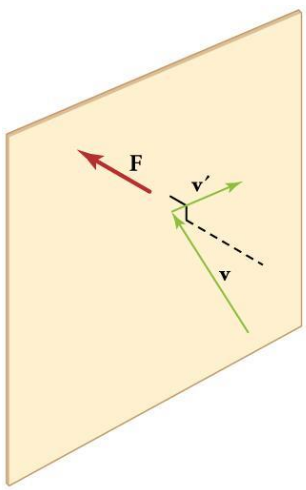
The ideal gas law can be considered to be another manifestation of the law of conservation of energy. Work done on a gas results in an increase in its energy, increasing pressure and/or temperature, or decreasing volume. This increased energy can also be viewed as increased internal kinetic energy, given the gas's atoms and molecules.

13.10 The Ideal Gas Law and Energy

Let us now examine the role of energy in the behavior of gases. When you inflate a bike tire by hand, you do work by repeatedly exerting a force through a distance. This energy goes into increasing the pressure of air inside the tire and increasing the temperature of the pump and the air.

The ideal gas law is closely related to energy: the units on both sides are joules. The right-hand side of the ideal gas law in $PV=NkT$ is NkT. This term is roughly the amount of translational kinetic energy of N atoms or molecules at an absolute temperature T. The left-hand side of the ideal gas law is PV, which also has the units of joules. We know from our study of fluids that pressure is one type of potential energy per unit volume, so pressure multiplied by volume is energy. The important point is that there is energy in a gas related to both its pressure and its volume. The energy can be changed when the gas is doing work as it expands — similar to what occurs in gasoline or steam engines and turbines.

We have developed macroscopic definitions of pressure and temperature. Pressure is the force divided by the area on which the force is exerted, and temperature is measured with a thermometer. We gain a better understanding of pressure and temperature from the kinetic theory of gases, which assumes that atoms and molecules are in continuous random motion.



3

Figure 13.11: When a molecule collides with a rigid wall, the component of its momentum perpendicular to the wall is reversed. A force is thus exerted on the wall, creating pressure.

Figure 13.11 shows an elastic collision of a gas molecule with the wall of a container, so that it exerts a force on the wall (by Newton's third law). Because a huge number of molecules will collide with the wall in a short time, we observe an average force per unit area. These collisions are the source of pressure in a gas. As the number of molecules increases, the number of collisions and thus the pressure increase. Similarly, the gas pressure is higher if the average velocity of molecules is higher. The actual relationship is as follows:

$$PV = (1/3)Nmv_{av}^2$$

where P is the pressure (average force per unit area), V is the volume of gas in the container, N is the number of molecules in the container, m is the mass of a molecule, and v_{av}^2 is the average of the molecular speed squared.

What can we learn from this atomic and molecular version of the ideal gas law? We can derive a relationship between temperature and the average translational kinetic energy of molecules in a gas. Recall the previous expression of the ideal gas law:

$$PV = NkT$$

Equating the right-hand side of this equation with the right-hand side of $PV = (1/3)Nmv_{av}^2$ gives

$$(1/3)Nmv_{av}^2 = NkT$$

Atomic and Molecular Origin of Pressure in a Gas

Figure 13.12 shows a box filled with a gas. We know from our previous discussions that putting more gas into the box produces greater pressure, and that increasing the temperature of the gas also produces greater pressure. But why should increasing the temperature of the gas increase the pressure in the box? A look at the atomic and molecular scale gives us some answers, and an alternative expression for the ideal gas law.

The figure shows an expanded view of an elastic collision of a gas molecule with the wall of a container. Calculating the average force exerted by such molecules will lead us to the ideal gas law, and to the connection between temperature and molecular kinetic energy. We assume that a molecule is small compared with the separation of molecules in the gas, and that its interaction with other molecules can be ignored. We also assume the wall is rigid and that the molecule's direction changes, but that its speed remains constant (and hence its kinetic energy and the magnitude of its momentum remain constant as well). This assumption is not always valid, but the same result is obtained with a more detailed description of the molecule's exchange of energy and momentum with the wall.

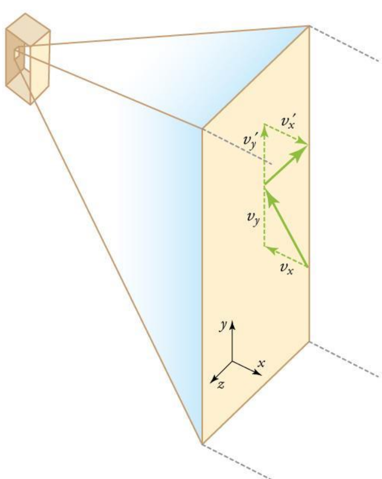


Figure 13.12: Gas in a box exerts an outward pressure on its walls. A molecule colliding with a rigid wall has the direction of its velocity and momentum in the *x*-direction reversed. This direction is perpendicular to the wall. The components of its velocity momentum in the *y*- and *z*-directions are not changed, which means there is no force parallel to the wall.

If the molecule's velocity changes in the *x*-direction, its momentum changes from $-mv_x$ to $+mv_x$. Thus, its change in momentum is $\Delta mv = +mv_x - (-mv_x) = 2mv_x$. The force exerted on the molecule is given by

$$F = \Delta p / \Delta t = (2mv_x) / \Delta t$$

There is no force between the wall and the molecule until the molecule hits the wall. During the short time of the collision, the force between the molecule and wall is relatively large. We are looking for an average force; we take Δt to be the average time between collisions of the molecule with this wall. It is the time it would take the molecule to go across the box and back (a distance $2l$) at a speed of v_x. Thus $\Delta t = 2l / v_x$, and the expression for the force becomes

$$F = (2mv_x) / (2l/v_x) = (mv^2_x) / l.$$

This force is due to *one* molecule. We multiply by the number of molecules *N* and use their average squared velocity to find the force,

$$F = N [(mv_{avx}^2) / l]$$

where the bar over a quantity means its average value. We would like to have the force in terms of the speed *v*, rather than the *x*-component of the velocity. We note that the total velocity squared is the sum of the squares of its components, so that

$$v_{av}^2 = v_{avx}^2 + v_{avy}^2 + v_{avz}^2$$

Because the velocities are random, their average components in all directions are the same:

$$v_{avx}^2 = v_{avy}^2 = v_{avz}^2$$

Thus,

$$v_{av}^2 = 3v_{avx}^2$$

or

$$v_{avx}^2 = (1/3)v_{av}^2$$

Substituting $1/3v_{av}^2$ into the expression for *F* gives

$$F = N [(mv_{av}^2) / 3l]$$

The pressure is F/A, so that we obtain

$$P = F/A = N\,[\,(mv_{\mathrm{av}}^2) / (3Al)\,] = 1/3\,[\,(Nmv_{\mathrm{av}}^2) / V]$$

where we used $V = Al$ for the volume. This gives the important result

$$PV = 1/3\,(Nmv_{\mathrm{av}}^2)$$

This equation is another expression of the ideal gas law.

We can get the average kinetic energy of a molecule, $½(mv^2)$, from the right-hand side of the equation by canceling N and multiplying by 3/2. This calculation produces the result that the average kinetic energy of a molecule is directly related to absolute temperature.

$$\mathrm{KE}_{\mathrm{av}} = ½(mv_{av}^2) = 3/2(kT)$$

The average translational kinetic energy of a molecule, $\mathrm{KE}_{\mathrm{av}}$ is called thermal energy**.** The equation $\mathrm{KE}_{\mathrm{av}} = ½(mv_{av}^2) = 3/2(kT)$ is a molecular interpretation of temperature, and it has been found to be valid for gases and reasonably accurate in liquids and solids. It is another definition of temperature based on an expression of the molecular energy.

It is sometimes useful to rearrange $\mathrm{KE}_{\mathrm{av}} = ½(mv_{av}^2) = 3/2(kT)$**,** and solve for the average speed of molecules in a gas in terms of temperature,

$$\sqrt{v_{av}^2} = v_{\mathrm{rms}} = \sqrt{[(3kT) / m]}$$

where $v$rms stands for root-mean-square (rms) speed.

13.11 Distribution of Molecular Speeds

The motion of molecules in a gas is random in magnitude and direction for individual molecules, but a gas of many molecules has a predictable distribution of molecular speeds. This distribution is called the *Maxwell-Boltzmann distribution*, after its originators, who calculated it based on kinetic theory, and has since been confirmed experimentally. (See Figure 13.13) The distribution has a long tail, because a few molecules may go several times the rms speed. The most probable speed vp is less than the rms speed v_{rms}. Figure 13.14 shows that the curve is shifted to higher speeds at higher temperatures, with a broader range of speeds.

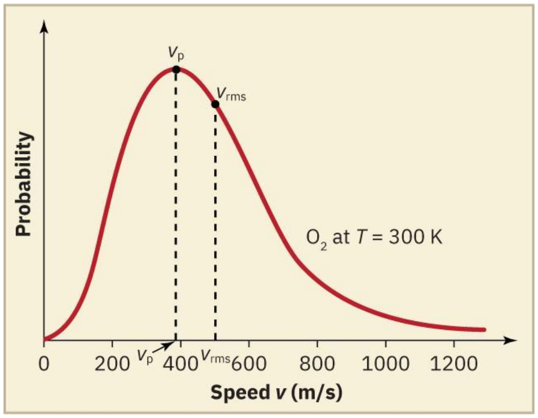


Figure 13.13: The Maxwell-Boltzmann distribution of molecular speeds in an ideal gas. The most likely speed v_p is less than the rms speed v_{rms}. Although very high speeds are possible, only a tiny fraction of the molecules have speeds that are an order of magnitude greater than *v*rms.

The distribution of thermal speeds depends strongly on temperature. As temperature increases, the speeds are shifted to higher values and the distribution is broadened.

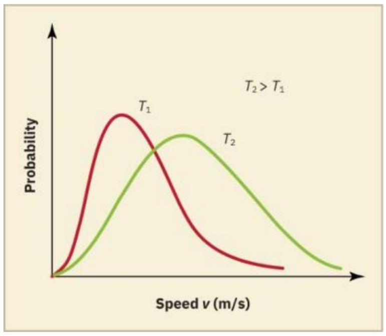


Figure 13.14: The Maxwell-Boltzmann distribution is shifted to higher speeds and is broadened at higher temperatures.

Up to now, we have considered the behavior of ideal gases. Real gases are like ideal gases at high temperatures. At lower temperatures, however, the interactions between the molecules and their volumes cannot be ignored. The molecules are very close (condensation occurs) and there is a dramatic decrease in volume, as seen in Figure 3.15. The substance changes from a gas to a liquid. When a liquid is cooled to even lower temperatures, it becomes a solid. The volume never reaches zero because of the finite volume of the molecules.

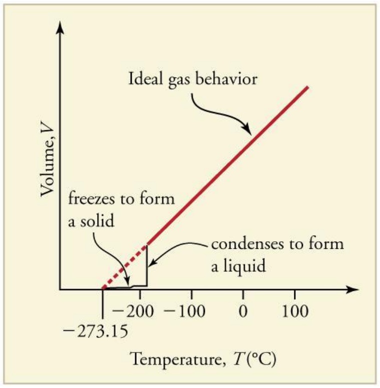


Figure 13.15: A sketch of volume versus temperature for a real gas at constant pressure. The linear (straight line) part of the graph represents ideal gas behavior—volume and temperature are directly and positively related and the line extrapolates to zero volume at –273.15ºC, or absolute zero. When the gas becomes a liquid, however, the volume actually decreases precipitously at the liquefaction point. The volume decreases slightly once the substance is solid, but it never becomes zero.

High pressure may also cause a gas to change phase to a liquid. Carbon dioxide, for example, is a gas at room temperature and atmospheric pressure, but becomes a liquid under sufficiently high pressure. If the pressure is reduced, the temperature drops and the liquid carbon dioxide solidifies into a snow-like substance at the temperature –78ºC. Solid CO_2 is called "dry ice." Another example of a gas that can be in a liquid phase is liquid nitrogen (LN_2). LN_2 is made by liquefaction of atmospheric air (through compression and cooling). It boils at 77 K (–196ºC) at atmospheric pressure. LN2 is useful as a refrigerant and allows for the preservation of blood, sperm, and other biological materials. It is also used to reduce noise in electronic sensors and equipment, and to help cool down their current-carrying wires. In dermatology, LN_2 is used to freeze and painlessly remove warts and other growths from the skin.

PV Diagrams

We can examine aspects of the behavior of a substance by plotting a graph of pressure versus volume, called a *PV* diagram. When the substance behaves like an ideal gas, the ideal gas law describes the relationship between its pressure and volume. That is,

$$PV = NkT \text{(ideal gas)}$$

Now, assuming the number of molecules and the temperature are fixed,

$$PV = \text{constant(ideal gas, constant temperature).}$$

For example, the volume of the gas will decrease as the pressure increases. If you plot the relationship PV = constant on a *PV* diagram, you find a hyperbola. Figure 13.16 shows a graph of pressure versus volume. The hyperbolas represent ideal-gas behavior at various fixed temperatures, and are called *isotherms*. At lower temperatures, the curves begin to look less like hyperbolas—the gas is not behaving ideally and may even contain liquid. There is a critical point—that is, a critical temperature—above which liquid cannot exist. At sufficiently high pressure above the critical point, the gas will have the density of a liquid but will not condense. Carbon dioxide, for example, cannot be liquefied at a temperature above 31.0ºC. Critical pressure is the minimum pressure needed for liquid to exist at the critical temperature.

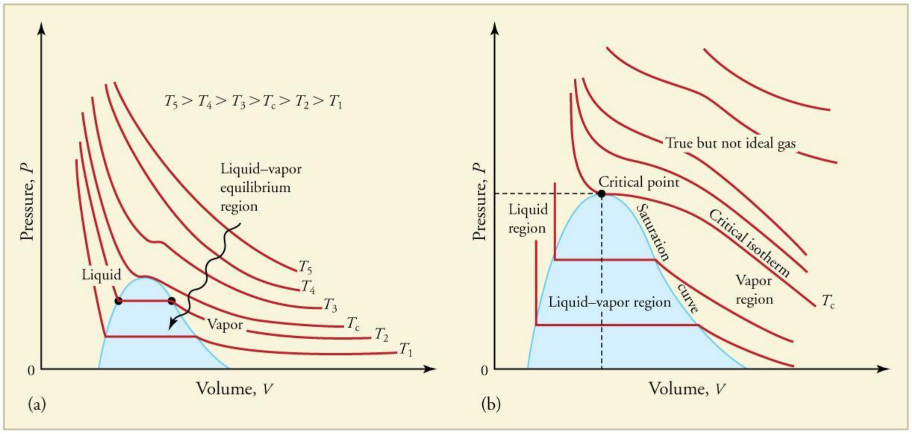


Figure 13.16: *PV* diagrams. (a) Each curve (isotherm) represents the relationship between *P* and *V* at a fixed temperature; the upper curves are at higher temperatures. The lower curves are not hyperbolas, because the gas is no longer an ideal gas. (b) An expanded portion of the *PV* diagram for low temperatures, where the phase can change from a gas to a liquid. The term "vapor" refers to the gas phase when it exists at a temperature below the boiling temperature.

Table 13.3 Critical Temperatures and Pressures

Substance	**Critical temperature**		**Critical pressure**	
	K	ºC	Pa	atm
Water	647.4	374.3	22.12×10^6	219.0
Sulfur dioxide	430.7	157.6	7.88×10^6	78.0
Ammonia	405.5	132.4	11.28×10^6	111.7
Carbon dioxide	304.2	31.1	7.39×10^6	73.2
Oxygen	154.8	−118.4	5.08×10^6	50.3

Substance	Critical temperature		Critical pressure	
	K	ºC	Pa	atm
Nitrogen	126.2	−146.9	3.39×10^6	33.6
Hydrogen	33.3	−239.9	1.30×10^6	12.9
Helium	5.3	−267.9	0.229×10^6	2.27

Table 13.3 Critical Temperatures and Pressures

13.12 Phase Diagrams

The plots of pressure versus temperatures provide considerable insight into thermal properties of substances. There are well-defined regions on these graphs that correspond to various phases of matter, so *PT* graphs are called phase diagrams. Figure 13.17 shows the phase diagram for water. Using the graph, if you know the pressure and temperature you can determine the phase of water. The solid lines—boundaries between phases—indicate temperatures and pressures at which the phases coexist (that is, they exist together in ratios, depending on pressure and temperature). For example, the boiling point of water is 100ºC at 1.00 atm. As the pressure increases, the boiling temperature rises steadily to 374ºC at a pressure of 218 atm. A pressure cooker (or even a covered pot) will cook food faster because the water can exist as a liquid at temperatures greater than 100ºC without all boiling away. The curve ends at a point called the *critical point*, because at higher temperatures the liquid phase does not exist at any pressure. The critical point occurs at the critical temperature, as you can see for water from Table 13.3. The critical temperature for oxygen is –118ºC, so oxygen cannot be liquefied above this temperature.

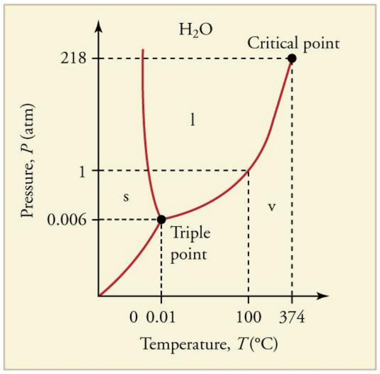


Figure 13.17: The phase diagram (*PT* graph) for water. Note that the axes are nonlinear and the graph is not to scale. This graph is simplified—there are several other exotic phases of ice at higher pressures.

Similarly, the curve between the solid and liquid regions in Figure 13.17 gives the melting temperature at various pressures. For example, the melting point is 0ºC at 1.00 atm, as expected. Note that, at a fixed temperature, you can change the phase from solid (ice) to liquid (water) by increasing the pressure. Ice melts from pressure in the hands of a snowball maker. From the phase diagram, we can also say that the melting temperature of ice falls with increased pressure. When a car is driven over snow, the increased pressure from the tires melts the snowflakes; afterwards the water refreezes and forms an ice layer.

At sufficiently low pressures there is no liquid phase, but the substance can exist as either gas or solid. For water, there is no liquid phase at pressures below 0.00600 atm. The phase change from solid to gas is called sublimation. It accounts for large losses of snowpack that never make it into a river, the routine automatic defrosting of a freezer, and the freeze-drying process applied to many foods. Carbon dioxide, on the other hand, sublimates at standard atmospheric pressure of 1 atm. (The solid form of CO_2 is known as dry ice because it does not melt. Instead, it moves directly from the solid to the gas state.)

All three curves on the phase diagram meet at a single point, the triple point, where all three phases exist in equilibrium. For water, the triple point occurs at 273.16 K (0.01ºC), and is a more accurate calibration temperature than the melting point of water at 1.00 atm, or 273.15 K (0.0ºC). See Table 13.4 for the triple point values of other substances.

Equilibrium

Liquid and gas phases are in equilibrium at the boiling temperature. (See Figure 13.18.) If a substance is in a closed container at the boiling point, then the liquid is boiling and the gas is condensing at the same rate without net change in their relative amount. Molecules in the liquid escape as a gas at the same rate at which gas molecules stick to the liquid, or form droplets and become part of the liquid phase. The combination of temperature and pressure has to be "just right"; if the temperature and pressure are increased, equilibrium is maintained by the same increase of boiling and condensation rates.

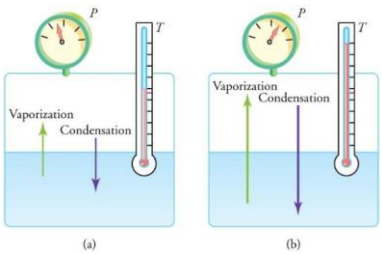


Figure 13.18: Equilibrium between liquid and gas at two different boiling points inside a closed container. (a) The rates of boiling and condensation are equal at this combination of temperature and pressure, so the liquid and gas phases are in equilibrium. (b) At a higher temperature, the boiling rate is faster and the rates at which molecules leave the liquid and enter the gas are also faster. Because there are more molecules in the gas, the gas pressure is higher and the rate at

which gas molecules condense and enter the liquid is faster. As a result, the gas and liquid are in equilibrium at this higher temperature.

Substance	Temperature		Pressure	
	K	ºC	Pa	atm
Water	273.16	0.01	6.10×10^2	0.00600
Carbon dioxide	216.55	−56.60	5.16×10^5	5.11
Sulfur dioxide	197.68	−75.47	1.67×10^3	0.0167
Ammonia	195.40	−77.75	6.06×10^3	0.0600
Nitrogen	63.18	−210.0	1.25×10^4	0.124
Oxygen	54.36	−218.8	1.52×10^2	0.00151
Hydrogen	13.84	−259.3	7.04×10^3	0.0697

Table 13.4: Triple Point Temperatures and Pressures

One example of equilibrium between liquid and gas is that of water and steam at 100ºC and 1.00 atm. This temperature is the boiling point at that pressure, so they should exist in equilibrium. Why does an open pot of water at 100ºC boil completely away? The gas surrounding an open pot is not pure water: it is mixed with air. If pure water and steam are in a closed container at 100ºC and 1.00 atm, they would coexist—but with air over the pot, there are fewer water molecules to condense, and water boils. What about water at 20.0ºC and 1.00 atm? This temperature and pressure correspond to the liquid region, yet an open glass of water at this temperature will completely evaporate. Again, the gas around it is air and not pure water vapor, so that the reduced evaporation rate is greater than the condensation rate of water from dry air. If the glass is sealed, then the liquid phase remains. We call the gas phase a vapor when it exists, as it does for water at 20.0ºC, at a temperature below the boiling temperature.

13.13 Vapor Pressure, Partial Pressure, and Dalton's Law

Vapor pressure is defined as the pressure at which a gas coexists with its solid or liquid phase. Vapor pressure is created by faster molecules that break away from the liquid or solid and enter the gas phase. The vapor pressure of a substance depends on both the substance and its temperature—an increase in temperature increases the vapor pressure.

Partial pressure is defined as the pressure a gas would create if it occupied the total volume available. In a mixture of gases, *the total pressure is the sum of partial pressures of the component gases*, assuming ideal gas behavior and no chemical reactions between the components. This law is known as Dalton's law of partial pressures, after the English scientist John Dalton (1766–1844), who proposed it. Dalton's law is based on kinetic theory, where each gas creates its pressure by molecular collisions, independent of other gases present. It is consistent with the fact that pressures add according to Pascal Principle. Thus, water evaporates and ice sublimates when their vapor pressures exceed the partial pressure of water vapor in the surrounding mixture of gases. If their vapor pressures are less than the partial pressure of water vapor in the surrounding gas, liquid droplets or ice crystals (frost) form.

When we say humidity, we really mean relative humidity. Relative humidity tells us how much water vapor is in the air compared with the maximum possible. At its maximum, denoted as saturation, the relative humidity is 100%, and evaporation is inhibited. The amount of water vapor in the air depends on temperature. For example, relative humidity rises in the evening, as air temperature declines, sometimes reaching the dew point. At the dew point temperature, relative humidity is 100%, and fog may result from the condensation of water droplets if they are small enough to stay in suspension. Conversely, if you wish to dry something (perhaps your hair), it is more effective to blow hot air over it rather than cold air, because, among other things, the increase in temperature increases the energy of the molecules, so the rate of evaporation increases.

The amount of water vapor in the air depends on the vapor pressure of water. The liquid and solid phases are continuously giving off vapor because some of the molecules have high enough speeds to enter the gas phase; see Figure 13.19 (a). If a lid is placed over the container, as in Figure 13.19 (b), evaporation continues, increasing the pressure, until sufficient vapor has built up for condensation to balance evaporation. Then equilibrium has been achieved, and the vapor pressure is equal to the partial pressure of water in the container. Vapor pressure increases with temperature because molecular speeds are higher as temperature increases. Table 13.5 gives representative values of water vapor pressure over a range of temperatures.

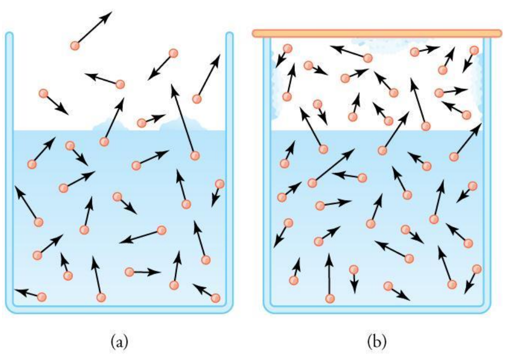


Figure 13.19: (a) Because of the distribution of speeds and kinetic energies, some water molecules can break away to the vapor phase even at temperatures below the ordinary boiling point. (b) If the container is sealed, evaporation will continue until there is enough vapor density for the condensation rate to equal the evaporation rate. This vapor density and the partial pressure it creates are the saturation values. They increase with temperature and are independent of the presence of other gases, such as air. They depend only on the vapor pressure of water.

Relative humidity is related to the partial pressure of water vapor in the air. At 100% humidity, the partial pressure is equal to the vapor pressure, and no more water can enter the vapor phase. If the partial pressure is less than the vapor pressure, then evaporation will take place, as humidity is less than 100%. If the partial pressure is greater than the vapor pressure, condensation takes place. In everyday language, people sometimes refer to the capacity of air to “hold” water vapor, but this is not actually what happens. The water vapor is not held by the air. The amount of water in air is determined by the vapor pressure of water and has nothing to do with the properties of air.

Temperature (°C)	**Vapor pressure (Pa)**	**Saturation vapor density (g/m^3)**
−50	4.0	0.039
−20	1.04×102	0.89
−10	2.60×102	2.36
0	6.10×102	4.84
5	8.68×102	6.80
10	1.19×103	9.40
15	1.69×103	12.8
20	2.33×103	17.2
25	3.17×103	23.0
30	4.24×103	30.4
37	6.31×103	44.0
40	7.34×103	51.1
50	1.23×104	82.4
60	1.99×104	130
70	3.12×104	197
80	4.73×104	294
90	7.01×104	418
95	8.59×104	505
100	**1.01×105**	**598**
120	1.99×105	1095
150	4.76×105	2430
200	1.55×106	7090
220	2.32×106	10,200

Table 13.5: Saturation Vapor Density of Water

Percent Relative Humidity

We define percent relative humidity as the ratio of vapor density to saturation vapor density, or

percent relative humidity = [(vapor density) / (saturation vapor density)]×100

For examples and answers, please refer to OpenStax.com questions and answers given on their website or to the College Physics 2e - https://openstax.org/details/books/college-physics-2e.

Chapter 14

14.0 Objectives

At the end of this lesson, students should be able to,

- Calculate heat, temperature, and heat capacity.
- Identify phase changes.
- Identify heat transfer methods.
- Compare and calculate conduction, convection, and radiation.

14.1 Introduction

This chapter covers heat, temperature, heat capacity, phase changes, heat transfer, conduction, convection, and radiation and related calculations.

We have already defined work as force times distance and learned that work done on an object changes its kinetic energy. We also realize that temperature is proportional to the (average) kinetic energy of atoms and molecules. We say that a thermal system has a certain internal energy: its internal energy is higher if the temperature is higher. If two objects at different temperatures are brought in contact with each other, energy is transferred from the hotter to the colder object until equilibrium is reached and the bodies reach thermal equilibrium (i.e., they are at the same temperature). No work is done by either object, because no force acts through a distance. The transfer of energy is caused by the temperature difference and ceases once the temperatures are equal. These observations lead to the following definition of heat: Heat is the spontaneous transfer of energy due to a temperature difference.

14.2 Heat

Heat is often confused with temperature. For example, we may say the heat was unbearable, when we actually mean that the temperature was high. Heat is a form of energy, whereas temperature is not. The misconception arises because we are sensitive to the flow of heat, rather than the temperature.

Owing to the fact that heat is a form of energy, it has the SI unit of *joule* (J). The *calorie* (cal) is a common unit of energy, defined as the energy needed to change the temperature of 1.00 g of water by 1.00ºC —specifically, between 14.5ºC and 15.5ºC, since there is a slight temperature dependence. Perhaps the most common unit of heat is the kilocalorie (kcal), which is the energy needed to change the temperature of 1.00 kg of water by 1.00ºC. Since mass is most often specified in kilograms, kilocalorie is commonly used. Food calories (given the notation Cal, and sometimes called “big calorie”) are actually kilocalories (1kilocalorie=1000 calories), a fact not easily determined from package labeling.

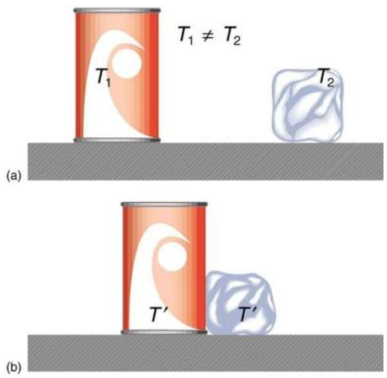


Figure 14.1: In figure (a) the soft drink and the ice have different temperatures, T_1 and T_2, and are not in thermal equilibrium. In figure (b), when the soft drink and ice are allowed to interact, energy is transferred until they reach the same temperature T', achieving equilibrium. Heat transfer occurs due to the difference in temperatures. In fact, since the soft drink and ice are both in contact with the surrounding air and bench, the equilibrium temperature will be the same for both.

Mechanical Equivalent of Heat

It is also possible to change the temperature of a substance by doing work. Work can transfer energy into or out of a system. This realization helped establish the fact that heat is a form of energy. James Prescott Joule (1818–1889) performed many experiments to establish the mechanical equivalent of heat—*the work needed to produce the same effects as heat transfer*. In terms of the units used for these two terms, the best modern value for this equivalence is

$$1.000\text{kcal} = 4186\text{J}$$

We consider this equation as the conversion between two different units of energy.

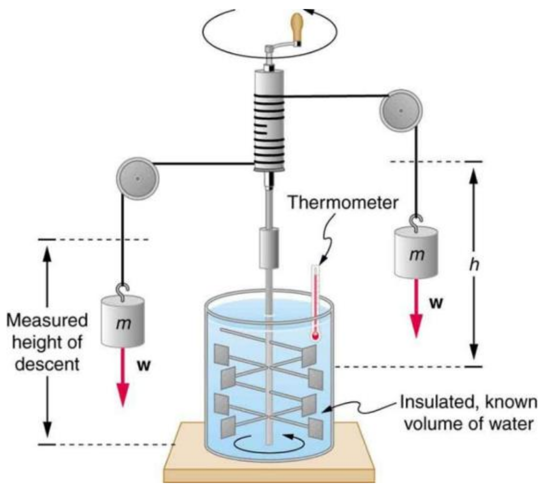


Figure 14.2: Schematic depiction of Joule's experiment that established the equivalence of heat and work.

The figure above shows one of Joule's most famous experimental setups for demonstrating the mechanical equivalent of heat. It demonstrated that work and heat can produce the same effects, and helped establish the principle of conservation of energy. Gravitational potential energy (PE) (work done by the gravitational force) is converted into kinetic energy (KE), and then randomized by viscosity and turbulence into increased average kinetic energy of atoms and molecules in the system, producing a temperature increase. His contributions to the field of thermodynamics were so significant that the SI unit of energy was named after him.

Heat added or removed from a system changes its internal energy and thus its temperature. Such a temperature increase is observed while cooking. However, adding heat does not necessarily increase the temperature. An example is melting of ice; that is, when a substance changes from one phase to another. Work done on the system or by the system can also change the internal energy of the system. Joule demonstrated that the temperature of a system can be increased by stirring. If an ice cube is rubbed against a rough surface, work is done by the frictional force. A system has a well-defined internal energy, but we cannot say that it has a certain "heat content" or "work content". We use the phrase "heat transfer" to emphasize its nature.

One of the major effects of heat transfer is temperature change: heating increases the temperature while cooling decreases it. We assume that there is no phase change and that no work is done on or by the system. Experiments show that the transferred heat depends on three factors—the change in temperature, the mass of the system, and the substance and phase of the substance.

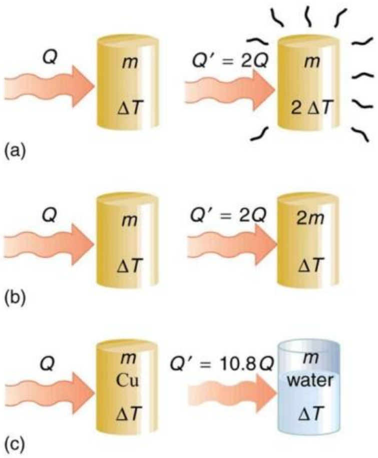


Figure 14.3: The heat Q transferred to cause a temperature change depends on the magnitude of the temperature change, the mass of the system, and the substance and phase involved. (a) The amount of heat transferred is directly proportional to the temperature change. To double the temperature change of a mass m, you need to add twice the heat. (b) The amount of heat transferred is also directly proportional to the mass. To cause an equivalent temperature change in a doubled mass, you need to add twice the heat. (c) The amount of heat transferred depends on the substance and its phase. If it takes an amount Q of heat to cause a temperature change ΔT in a given mass of copper, it will take 10.8 times that amount of heat to cause the equivalent temperature change in the same mass of water assuming no phase change in either substance.

The dependence on temperature change and mass are easily understood. Owing to the fact that the (average) kinetic energy of an atom or molecule is proportional to the absolute temperature, the internal energy of a system is proportional to the absolute temperature and the number of atoms or molecules. Owing to the fact that the transferred heat is equal to the change in the internal energy, the heat is proportional to the mass of the substance and the temperature change. The transferred heat also depends on the substance so that, for example, the heat necessary to raise the temperature is less for alcohol than for water. For the same substance, the transferred heat also depends on the phase (gas, liquid, or solid).

14.3 Heat Transfer and Temperature Change

The quantitative relationship between heat transfer and temperature change contains all three factors:

$$Q = mc\Delta T$$

where Q is the symbol for heat transfer, m is the mass of the substance, and ΔT is the change in temperature. The symbol c stands for specific heat and depends on the material and phase. The specific heat is the amount of heat necessary to change the temperature of 1.00 kg of mass by 1.00ºC. The specific heat c is a property of the substance; its SI unit is J/(kg·K) or J/(kg·ºC).

Recall that the temperature change (ΔT) is the same in units of kelvin and degrees Celsius. If heat transfer is measured in kilocalories, then *the unit of specific heat* is kcal/(kg·ºC).

Values of specific heat must generally be looked up in tables, because there is no simple way to calculate them. In general, the specific heat also depends on the temperature. Table 14.1 lists representative values of specific heat for various substances. Except for gases, the temperature and volume dependence of the specific heat of most substances is weak. We see from this table that the specific heat of water is five times that of glass and ten times that of iron, which means that it takes five times as much heat to raise the temperature of water the same amount as for glass and ten times as much heat to raise the temperature of water as for iron. In fact, water has one of the largest specific heats of any material, which is important for sustaining life on Earth.

Example - Calculating the Required Heat: Heating Water in an Aluminum Pan

A 0.500 kg aluminum pan on a stove is used to heat 0.250 liters of water from 20.0ºC to 80.0ºC. (a) How much heat is required? What percentage of the heat is used to raise the temperature of (b) the pan and (c) the water?

Strategy

The pan and the water are always at the same temperature. When you put the pan on the stove, the temperature of the water and the pan is increased by the same amount. We use the equation for the heat transfer for the given temperature change and mass of water and aluminum. The specific heat values for water and aluminum are given in Table 14.1.

Solution

Because water is in thermal contact with the aluminum, the pan and the water are at the same temperature.

1. Calculate the temperature difference:

$$\Delta T = T_\mathrm{f} - T_\mathrm{i} = 60.0^\circ\mathrm{C}$$

2. Calculate the mass of water. Because the density of water is $1000\mathrm{kg/m^3}$, one liter of water has a mass of 1 kg, and the mass of 0.250 liters of water is $m\mathrm{w}=0.250\mathrm{kg}$.
3. Calculate the heat transferred to the water. Use the specific heat of water in Table 14.1.:

$$Q_\mathrm{w} = m_\mathrm{w}c_\mathrm{w}\Delta T = (0.250\mathrm{kg})(4186\mathrm{J/kg^\circ C})(60.0^\circ\mathrm{C})=62.8\ \mathrm{kJ}$$

4. Calculate the heat transferred to the aluminum. Use the specific heat for aluminum in Table 14.1:

$$Q_\mathrm{Al} = m_\mathrm{Al}c_\mathrm{Al}\Delta\ T= (0.500\ \mathrm{kg})(900\ \mathrm{J/kg^\circ C})(60.0^\circ\mathrm{C})= 27.0 \times 104\mathrm{J} = 27.0\ \mathrm{kJ}$$

5. Compare the percentage of heat going into the pan versus that going into the water. First, find the total transferred heat:

$$Q_{\text{Total}} = Q_{\text{W}} + Q_{\text{Al}} = 62.8\text{kJ}+27.0\text{kJ} = 89.8\text{kJ}$$

Thus, the amount of heat going into heating the pan is

$$(27.0\text{kJ} / 89.8\text{kJ}) \times 100\% = 30.1\%$$

and the amount going into heating the water is

$$(62.8\text{kJ} / 89.8\text{kJ}) \times 100\% = 69.9\%$$

Discussion

In this example, the heat transferred to the container is a significant fraction of the total transferred heat. Although the mass of the pan is twice that of the water, the specific heat of water is over four times greater than that of aluminum. Therefore, it takes a bit more than twice the heat to achieve the given temperature change for the water as compared to the aluminum pan.

Figure 14.4: The smoking brakes on this truck are a visible evidence of the mechanical equivalent of heat.

Example - Calculating the Temperature Increase from the Work Done on a Substance: Truck Brakes Overheat on Downhill Runs

Truck brakes are used to control speed on a downhill run do work, converting gravitational potential energy into increased internal energy (higher temperature) of the brake material. This conversion prevents the gravitational potential energy from being converted into kinetic energy of the truck. The problem is that the mass of the truck is large compared with that of the brake material absorbing the energy, and the temperature increase may occur too fast for sufficient heat to transfer from the brakes to the environment.

Calculate the temperature increase of 100 kg of brake material with an average specific heat of 800 J(kg°C) if the material retains 10% of the energy from a 10,000-kg truck descending 75.0 m (in vertical displacement) at a constant speed.

Strategy

If the brakes are not applied, gravitational potential energy is converted into kinetic energy. When brakes are applied, gravitational potential energy is converted into internal energy of the brake material. We first calculate the gravitational potential energy (*Mgh*) that the entire truck loses in its descent and then find the temperature increase produced in the brake material alone.

Solution

1. Calculate the change in gravitational potential energy as the truck goes downhill

$$Mgh = (10{,}000 \text{ kg})\ (9.80 \text{ m/s}^2)(75.0 \text{ m}) = 7.35\times10^5 \text{J}.$$

2. Calculate the temperature from the heat transferred using $Q = Mgh$ and

$$\Delta\ T = Q\ /\ (mc)$$

where m is the mass of the brake material. Insert the values m = 100 kg and c = 800 J/(kg°C) to find

$$\Delta T = (7.35\times10^6 \text{J})\ /\ (100 \text{ kg})(800 \text{ J}(/\text{kg°C})) = 9.2\text{ºC}.$$

Discussion

This same idea underlies the recent hybrid technology of cars, where mechanical energy (gravitational potential energy) is converted by the brakes into electrical energy (battery).

Substances	**Specific heat (*c*)**	
Solids	J(kg°C)	kcal(kg°C)[1]
Aluminum	900	0.215
Asbestos	800	0.19
Concrete, granite (average)	840	0.20
Copper	387	0.0924
Glass	840	0.20
Gold	129	0.0308
Human body (average at 37 °C)	3500	0.83
Ice (average, -50°C to 0°C)	2090	0.50
Iron, steel	452	0.108
Lead	128	0.0305
Silver	235	0.0562
Wood	1700	0.4

Substances	Specific heat (c)	
Liquids		
Benzene	1740	0.415
Ethanol	2450	0.586
Glycerin	2410	0.576
Mercury	139	0.0333
Water (15.0 °C)	4186	1.000
Gases [2]	$Cv(Cp)$	$Cv(Cp)$
Air (dry)	721 (1015)	0.172 (0.242)
Ammonia	1670 (2190)	0.399 (0.523)
Carbon dioxide	638 (833)	0.152 (0.199)
Nitrogen	739 (1040)	0.177 (0.248)
Oxygen	651 (913)	0.156 (0.218)
Steam (100°C)	1520 (2020)	0.363 (0.482)

Table 14.1: Specific Heats[3] of Various Substances

Note that the previous example is an illustration of the mechanical equivalent of heat. Alternatively, the temperature increase could be produced by a blow torch instead of mechanically.

Example - Calculating the Final Temperature When Heat Is Transferred Between Two Bodies: Pouring Cold Water in a Hot Pan

Suppose you pour 0.250 kg of 20.0ºC water (about a cup) into a 0.500-kg aluminum pan off the stove with a temperature of 150ºC. Assume that the pan is placed on an insulated pad and that a negligible amount of water boils off. What is the temperature when the water and pan reach thermal equilibrium a short time later?

Strategy

The pan is placed on an insulated pad so that little heat transfer occurs with the surroundings. Originally the pan and water are not in thermal equilibrium: the pan is at a higher temperature than the water. Heat transfer then restores thermal equilibrium once the water and pan are in contact. Because heat transfer between the pan and water takes place rapidly, the mass of evaporated water is negligible and the magnitude of the heat lost by the pan is equal to the heat gained by the water. The exchange of heat stops once a thermal equilibrium between the pan and the water is achieved. The heat exchange can be written as $|Q_{hot}|=Q_{cold}$.

Solution

1. Use the equation for heat transfer $Q = mc\Delta T$ to express the heat lost by the aluminum pan in terms of the mass of the pan, the specific heat of aluminum, the initial temperature of the pan, and the final temperature:

$$Q_{hot} = m_{Al}c_{Al}\,(T_f - 150ºC)$$

2. Express the heat gained by the water in terms of the mass of the water, the specific heat of water, the initial temperature of the water and the final temperature:

$$Q_{cold} = m_W c_W (T_f - 20.0ºC)$$

3. Note that $Q_{hot}<0$ and $Q_{cold}>0$ and that they must sum to zero because the heat lost by the hot pan must be the same as the heat gained by the cold water:

$$Q_{cold} + Q_{hot} = 0$$

$$Q_{cold} = - Q_{hot}$$

$$m_W c_W (T_f - 20.0ºC) = - m_{Al} c_{Al} (T_f - 150ºC)$$

4. Bring all terms involving T_f on the left-hand side and all other terms on the right-hand side. Solve for T_f,

$$T_f = [m_{Al} c_{Al} (150ºC) + m_W c_W (20.0ºC)] / (m_{Al} c_{Al} + m_W c_W)$$

and insert the numerical values:

$$T_f = [(0.500 \text{ kg})(900 \text{ J}(/\text{kg}°\text{C}))(150ºC) + (0.250 \text{ kg})(4186 \text{ J}(/\text{kg}°\text{C}))(20.0ºC)] \;/$$

$$[(0.500 \text{ kg})(900 \text{ J}(/\text{kg}°\text{C})) + (0.250 \text{ kg})(4186 \text{ J}(/\text{kg}°\text{C}))]$$

$$= 88430 \text{ J} / 1496.5 \text{ J/ºC}$$

$$= 59.1ºC.$$

Discussion

This is a typical *calorimetry* problem—two bodies at different temperatures are brought in contact with each other and exchange heat until a common temperature is reached. Why is the final temperature so much closer to 20.0ºC than 150ºC? The reason is that water has a greater specific heat than most common substances and thus undergoes a small temperature change for a given heat transfer. A large body of water, such as a lake, requires a large amount of heat to increase its temperature appreciably. This explains why the temperature of a lake stays relatively constant during a day even when the temperature change of the air is large. However, the water temperature does change over longer times (e.g., summer to winter).

So far we have discussed temperature change due to heat transfer. No temperature change occurs from heat transfer if ice melts and becomes liquid water (i.e., during a phase change). For example, consider water dripping from icicles melting on a roof warmed by the Sun. Conversely, water freezes in an ice tray cooled by lower-temperature surroundings.

Energy is required to melt a solid because the cohesive bonds between the molecules in the solid must be broken apart such that, in the liquid, the molecules can move around at comparable kinetic energies; thus, there is no rise in temperature. Similarly, energy is needed to vaporize a liquid, because molecules in a liquid interact with each other via attractive forces. There is no temperature change until a phase change is complete. The temperature of a cup of soda initially at 0ºC stays at 0ºC until all the ice has melted. Conversely, energy is released during freezing and condensation, usually in the form of thermal energy. Work is done by cohesive forces when molecules are brought together. The corresponding energy must be given off (dissipated) to allow them to stay together Figure 14.5.

14.4 Phase Change

The energy involved in a phase change depends on two major factors: the number and strength of bonds or force pairs. The number of bonds is proportional to the number of molecules and thus to the mass of the sample. The strength of forces depends on the type of molecules. The heat *Q* required to change the phase of a sample of mass *m* is given by

$$Q=mL_f \text{ (melting/freezing)}$$

$$Q=mL_v \text{ (vaporization/condensation)}$$

where the latent heat of fusion, Lf, and latent heat of vaporization, Lv, are material constants that are determined experimentally. See (Table 14.2).

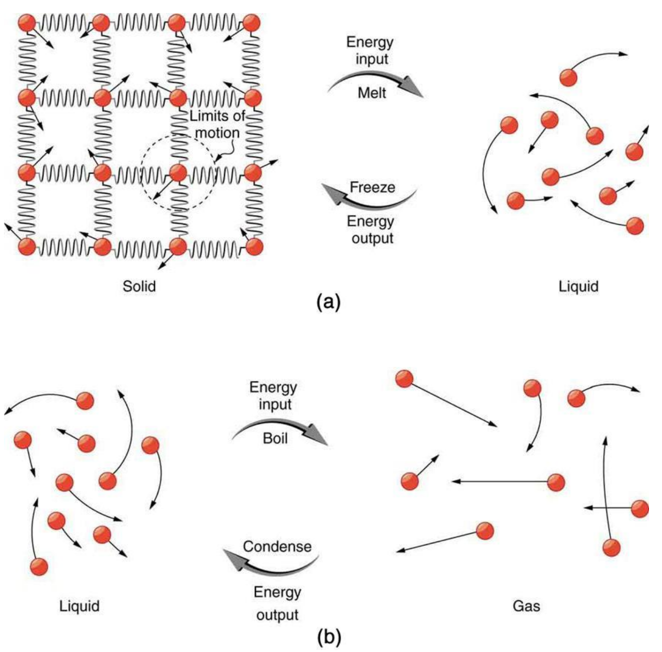


Figure 14.5: (a) Energy is required to partially overcome the attractive forces between molecules in a solid to form a liquid. That same energy must be removed for freezing to take place. (b) Molecules are separated by large distances when going from liquid to vapor, requiring significant energy to overcome molecular attraction. The same energy must be removed for condensation to take place. There is no temperature change until a phase change is complete.

Latent heat is measured in units of J/kg. Both L_f and Lv depend on the substance, particularly on the strength of its molecular forces as noted earlier. L_f and L_v are collectively called latent heat coefficients. They are *latent*, or hidden, because in phase changes, energy enters or leaves a system without causing a temperature change in the system; so, in effect, the energy is hidden. Table 14.2 lists representative values of L_f and L_v, together with melting and boiling points.

The table shows that significant amounts of energy are involved in phase changes. Let us look, for example, at how much energy is needed to melt a kilogram of ice at 0ºC to produce a kilogram of water at 0°C. Using the equation for a change in temperature and the value for water from Table 14.2, we find that $Q = mL_f = (1.0\text{kg})(334\text{kJ/kg}) = 334\text{kJ}$ is the energy to melt a kilogram of ice. This is a lot of energy as it represents the same amount of energy needed to raise the temperature of 1 kg of liquid water from 0ºC to 79.8ºC. Even more energy is required to vaporize water; it would take 2256 kJ to change 1 kg of liquid water at the normal boiling point (100ºC at atmospheric pressure) to steam (water vapor). This example shows that the energy for a phase change is enormous compared to energy associated with temperature changes without a phase change.

		L_f			***L_v***	
Substance	**Melting point (°C)**	**kJ/kg**	**kcal/kg**	**Boiling point (°C)**	**kJ/kg**	**kcal/kg**
Helium	−269.7	5.23	1.25	−268.9	20.9	4.99
Hydrogen	−259.3	58.6	14.0	−252.9	452	108
Nitrogen	−210.0	25.5	6.09	−195.8	201	48.0
Oxygen	−218.8	13.8	3.30	−183.0	213	50.9
Ethanol	−114	104	24.9	78.3	854	204
Ammonia	−75		108	−33.4	1370	327
Mercury	−38.9	11.8	2.82	357	272	65.0
Water	0.00	334	79.8	100.0	2256 [4]	540[5]
Water				37		580
Sulfur	119	38.1	9.10	444.6	326	77.9
Lead	327	24.5	5.85	1750	871	208
Antimony	631	165	39.4	1440	561	134
Aluminum	660	380	90	2450	11400	2720
Silver	961	88.3	21.1	2193	2336	558
Gold	1063	64.5	15.4	2660	1578	377
Copper	1083	134	32.0	2595	5069	1211
Uranium	1133	84	20	3900	1900	454
Tungsten	3410	184	44	5900	4810	1150

Table 14.2: Heats of Fusion and Vaporization

Phase changes can have a tremendous stabilizing effect even on temperatures that are not near the melting and boiling points, because evaporation and condensation (conversion of a gas into a liquid state) occur even at temperatures below the boiling point. Take, for example, the fact that air temperatures in humid climates rarely go above 35.0ºC, which is because most heat transfer goes into evaporating water into the air. Similarly, temperatures in humid weather rarely fall below the dew point because enormous heat is released when water vapor condenses.

We examine the effects of phase change more precisely by considering adding heat into a sample of ice at −20ºC (Figure 14.6). The temperature of the ice rises linearly, absorbing heat at a constant rate of 0.50 cal/g·ºC until it reaches 0ºC. Once at this temperature, the ice begins to melt until all the ice has melted, absorbing 79.8 cal/g of heat. The temperature remains constant at 0ºC during this phase change. Once all the ice has melted, the temperature of the liquid water rises, absorbing heat at a new constant rate of 1.00 cal/g·ºC. At 100ºC, the water begins to boil and the temperature again remains constant while the water absorbs 539 cal/g of heat during this phase change. When all the liquid has become steam vapor, the temperature rises again, absorbing heat at a rate of 0.482 cal/g·ºC.

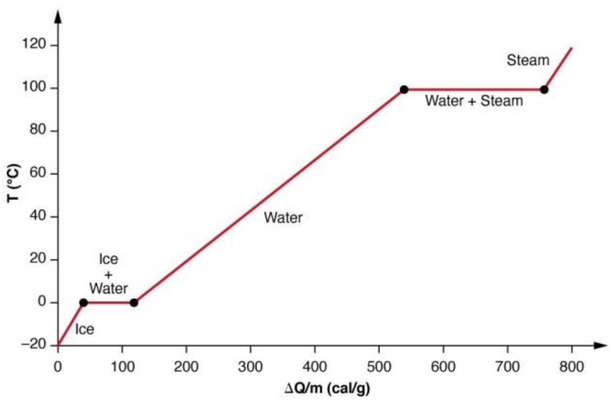


Figure 14.6: A graph of temperature versus energy added. The system is constructed so that no vapor evaporates while ice warms to become liquid water, and so that, when vaporization occurs, the vapor remains in of the system. The long stretches of constant temperature values at 0ºC and 100ºC reflect the large latent heat of melting and vaporization, respectively.

Water can evaporate at temperatures below the boiling point. More energy is required than at the boiling point, because the kinetic energy of water molecules at temperatures below 100ºC is less than that at 100ºC, hence less energy is available from random thermal motions. Take, for example, the fact that, at body temperature, perspiration from the skin requires a heat input of 2428 kJ/kg, which is about 10 percent higher than the latent heat of vaporization at 100ºC. This heat comes from the skin, and thus provides an effective cooling mechanism in hot weather.

High humidity inhibits evaporation, so that body temperature might rise, leaving unevaporated sweat on your brow.

Example - Calculate Final Temperature from Phase Change: Cooling Soda with Ice Cubes

Three ice cubes are used to chill a soda at 20ºC with mass m_{soda}=0.25 kg. The ice is at 0ºC and each ice cube has a mass of 6.0 g. Assume that the soda is kept in a foam container so that heat loss can be ignored. Assume the soda has the same heat capacity as water. Find the final temperature when all ice has melted.

Strategy

The ice cubes are at a melting temperature of 0ºC. Heat is transferred from the soda to the ice for melting. Melting of ice occurs in two steps: first the phase change occurs and solid (ice) transforms into liquid water at the melting temperature, then the temperature of this water rises. Melting yields water at 0ºC, so more heat is transferred from the soda to this water until the water plus soda system reaches thermal equilibrium,

$$Q_{ice} = -Q_{soda}$$

The heat transferred to the ice is $Q_{ice} = m_{ice}L_f + m_{ice}c_W(T_f - 0ºC)$. The heat given off by the soda is $Q_{soda} = m_{soda}c_W(T_f - 20ºC)$. Since no heat is lost, $Q_{ice} = -Q_{soda}$, so that

$$m_{ice}L_f + m_{ice}c_W(T_f - 0ºC) = -m_{soda}c_W(T_f - 20ºC)$$

Bring all terms involving T_f on the left-hand-side and all other terms on the right-hand-side. Solve for the unknown quantity T_f:

$$T_f = m_{soda}c_W(20ºC) - m_{ice}L_f(m_{soda}+m_{ice})c_W$$

Solution

1. Identify the known quantities. The mass of ice is $m_{ice} = 3 \times 6.0\text{ g} = 0.018\text{ kg}$ and the mass of soda is m_{soda}=0.25 kg.
2. Calculate the terms in the numerator:

$$m_{soda}c_W(20ºC) = (0.25\text{ kg})(4186\text{ J}(/\text{kg}°\text{C}))(20ºC) = 20{,}930\text{ J}$$

and

$$m_{ice}L_f = (0.018\text{ kg})(334{,}000\text{ J/kg}) = 6012\text{ J}$$

3. Calculate the denominator:

$$(m_{soda} + m_{ice})c_W = (0.25\text{ kg} + 0.018\text{ kg})(4186\text{ J}/(\text{kg}\cdot ºC) = 1122\text{ J}/ºC.$$

4. Calculate the final temperature:

$$T_f = (20{,}930\ \text{J} - 6012\ \text{J}) / (1122\ \text{J/ºC}) = 13\text{ºC}$$

Discussion

This example illustrates the enormous energies involved during a phase change. The mass of ice is about 7 percent the mass of water but leads to a noticeable change in the temperature of soda. Although we assumed that the ice was at the freezing temperature, this is incorrect: the typical temperature is −6ºC.

We have seen that vaporization requires heat transfer to a liquid from the surroundings, so that energy is released by the surroundings. Condensation is the reverse process, increasing the temperature of the surroundings. This increase may seem surprising, since we associate condensation with cold objects—the glass in the figure, for example. However, energy must be removed from the condensing molecules to make a vapor condense. The energy is exactly the same as that required to make the phase change in the other direction, from liquid to vapor, and so it can be calculated from $Q = mL_v$.

Sublimation is the transition from solid to vapor phase. You may have noticed that snow can disappear into thin air without a trace of liquid water, or the disappearance of ice cubes in a freezer. The reverse is also true: Frost can form on very cold windows without going through the liquid stage. A popular effect is the making of "smoke" from dry ice, which is solid carbon dioxide. Sublimation occurs because the equilibrium vapor pressure of solids is not zero. Certain air fresheners use the sublimation of a solid to inject a perfume into the room. Moth balls are a slightly toxic example of a phenol (an organic compound) that sublimates, while some solids, such as osmium tetroxide, are so toxic that they must be kept in sealed containers to prevent human exposure to their sublimation-produced vapors.

All phase transitions involve heat. In the case of direct solid-vapor transitions, the energy required is given by the equation $Q = mL_s$, where L_s is the heat of sublimation, which is the energy required to change 1.00 kg of a substance from the solid phase to the vapor phase. L_s is analogous to L_f and L_v, and its value depends on the substance. Sublimation requires energy input, so that dry ice is an effective coolant, whereas the reverse process (i.e., frosting) releases energy. The amount of energy required for sublimation is of the same order of magnitude as that for other phase transitions.

The material presented in this section and the preceding section allows us to calculate any number of effects related to temperature and phase change. In each case, it is necessary to identify which temperature and phase changes are taking place and then to apply the appropriate equation. Keep in mind that heat transfer and work can cause both temperature and phase changes.

14.5 Heat Transfer – Conduction, Convection, and Radiation

Equally as interesting as the effects of heat transfer on a system are the methods by which this occurs. Whenever there is a temperature difference, heat transfer occurs. Heat transfer may occur rapidly, such as through a cooking pan, or slowly, such as through the walls of a picnic ice chest. We can control rates of heat transfer by choosing materials (such as thick wool clothing for the winter), controlling air movement (such as the use of weather stripping around doors), or by choice of color (such as a white roof to reflect summer sunlight). So many processes involve heat transfer, so that it is hard to imagine a situation where no heat transfer occurs. Yet every process involving heat transfer takes place by only three methods:

1. Conduction is heat transfer through stationary matter by physical contact. (The matter is stationary on a macroscopic scale—we know there is thermal motion of the atoms and molecules at any temperature above absolute zero.) Heat transferred between the electric burner of a stove and the bottom of a pan is transferred by conduction.
2. Convection is the heat transfer by the macroscopic movement of a fluid. This type of transfer takes place in a forced-air furnace and in weather systems, for example.
3. Heat transfer by radiation occurs when microwaves, infrared radiation, visible light, or another form of electromagnetic radiation is emitted or absorbed. An obvious example is the warming of the Earth by the Sun. A less obvious example is thermal radiation from the human body.

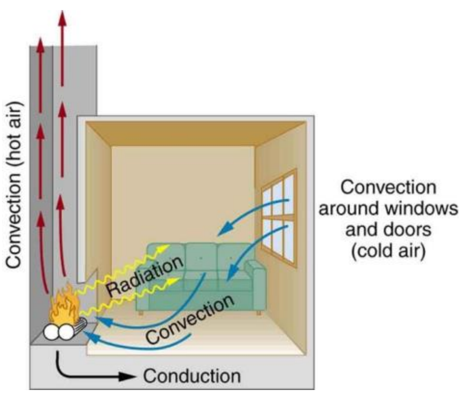


Figure 14.7: In a fireplace, heat transfer occurs by all three methods: conduction, convection, and radiation. Radiation is responsible for most of the heat transferred into the room. Heat transfer also occurs through conduction into the room, but at a much slower rate. Heat transfer by convection also occurs through cold air entering the room around windows and hot air leaving the room by rising up the chimney.

We examine these methods in some detail in the three following modules. Each method has unique and interesting characteristics, but all three do have one thing in common: they transfer heat solely because of a temperature difference Figure 14.7.

Figure 14.8: Insulation is used to limit the conduction of heat from the inside to the outside (in winters) and from the outside to the inside (in summers). (credit: Giles Douglas)

Your feet feel cold as you walk barefoot across the living room carpet in your cold house and then step onto the kitchen tile floor. This result is intriguing, since the carpet and tile floor are both at the same temperature. The different sensation you feel is explained by the different rates of heat transfer: the heat loss during the same time interval is greater for skin in contact with the tiles than with the carpet, so the temperature drop is greater on the tiles.

Some materials conduct thermal energy faster than others. In general, good conductors of electricity (metals like copper, aluminum, gold, and silver) are also good heat conductors, whereas insulators of electricity (wood, plastic, and rubber) are poor heat conductors. Figure 14.9 shows molecules in two bodies at different temperatures. The (average) kinetic energy of a molecule in the hot body is higher than in the colder body. If two molecules collide, an energy transfer from the molecule with greater kinetic energy to the molecule with less kinetic energy occurs. The cumulative effect from all collisions results in a net flux of heat from the hot body to the colder body. The heat flux thus depends on the temperature difference $\Delta T = T_{\text{hot}} - T_{\text{cold}}$. Therefore, you will get a more severe burn from boiling water than from hot tap water. Conversely, if the temperatures are the same, the net heat transfer rate falls to zero, and equilibrium is achieved. Owing to the fact that the number of collisions increases with increasing area, heat conduction depends on the cross-sectional area. If you touch a cold wall with your palm, your hand cools faster than if you just touch it with your fingertip.

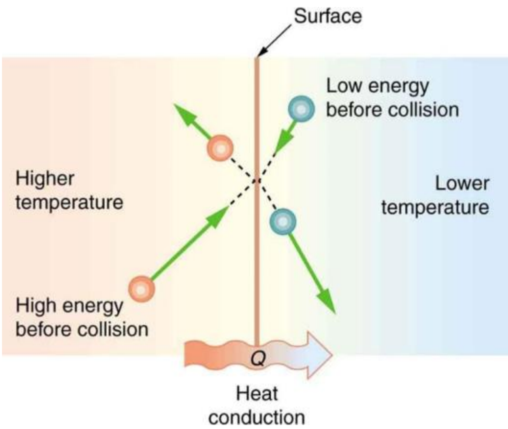


Figure 14.9: The molecules in two bodies at different temperatures have different average kinetic energies. Collisions occurring at the contact surface tend to transfer energy from high-temperature regions to low-temperature regions. In this illustration, a molecule in the lower temperature region (right side) has low energy before collision, but its energy increases after colliding with the contact surface. In contrast, a molecule in the higher temperature region (left side) has high energy before collision, but its energy decreases after colliding with the contact surface.

A third factor in the mechanism of conduction is the thickness of the material through which heat transfers. The figure below shows a slab of material with different temperatures on either side. Suppose that *T*2 is greater than *T*1, so that heat is transferred from left to right. Heat transfer from the left side to the right side is accomplished by a series of molecular collisions. The thicker the material, the more time it takes to transfer the same amount of heat. This model explains why thick clothing is warmer than thin clothing in winters, and why Arctic mammals protect themselves with thick blubber.

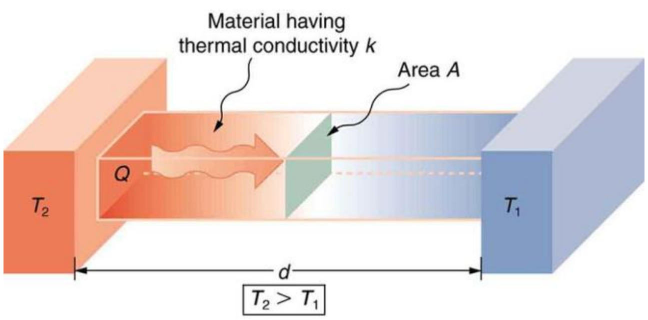


Figure 14.10: Heat conduction occurs through any material, represented here by a rectangular bar, whether window glass or walrus blubber. The temperature of the material is *T*2 on the left and *T*1 on the right, where T_2 is greater than T_1. The rate of heat transfer by conduction is

directly proportional to the surface area *A*, the temperature difference $T_2 - T_1$, and the substance's conductivity *k*. The rate of heat transfer is inversely proportional to the thickness *d*.

Lastly, the heat transfer rate depends on the material properties described by the coefficient of thermal conductivity. All four factors are included in a simple equation that was deduced from and is confirmed by experiments. The rate of conductive heat transfer through a slab of material, such as the one in Figure 14.10, is given by

$$Q/t = [kA(T2 - T1)] / d,$$

where Q/t is the rate of heat transfer in watts or kilocalories per second, *k* is the thermal conductivity of the material, *A* and *d* are its surface area and thickness, as shown in Figure 14.10, and $(T_2 - T_1)$ is the temperature difference across the slab. Table 14.3 gives representative values of thermal conductivity.

Example - Calculating Heat Transfer Through Conduction: Conduction Rate Through an Ice Box

A Styrofoam ice box has a total area of 0.950 m^2 and walls with an average thickness of 2.50 cm. The box contains ice, water, and canned beverages at 0ºC. The inside of the box is kept cold by melting ice. How much ice melts in one day if the ice box is kept in the trunk of a car at 35.0ºC?

Strategy

This question involves both heat for a phase change (melting of ice) and the transfer of heat by conduction. To find the amount of ice melted, we must find the net heat transferred. This value can be obtained by calculating the rate of heat transfer by conduction and multiplying by time.

Solution

1. Identify the knowns.

$$A = 0.950 m^2$$

$$d = 2.50 cm = 0.0250 \text{ m}$$

$$T_1 = 0ºC$$

$$T_2 = 35.0ºC, t = 1 \text{ day} = 24 \text{ hours} = 86{,}400 \text{ s}$$

2. Identify the unknowns. We need to solve for the mass of the ice, *m*. We will also need to solve for the net heat transferred to melt the ice, *Q*.
3. Determine which equations to use. The rate of heat transfer by conduction is given by

$$Q/t = [kA(T2 - T1)] / d$$

4. The heat is used to melt the ice: $Q = mL$f.
5. Insert the known values:

$$Q/t = [(0.010\ \text{J/s·m·°C})(0.950\ \text{m}^2)(35.0°\text{C} - 0°\text{C})\ ]\ /\ 0.0250\ \text{m} = 13.3\ \text{J/s}.$$

6. Multiply the rate of heat transfer by the time (1 day = 86,400 s):

$$Q = (Q/t)t = (13.3\ \text{J/s})(86{,}400\ \text{s}) = 1.15\times10^6\ \text{J}.$$

7. Set this equal to the heat transferred to melt the ice: $Q = mL_f$. Solve for the mass m:

$$M = Q\ /\ L_f = 1.15\times10^6\ \text{J}\ /\ 334\times10^3\ \text{J/kg} = 3.44\text{kg}.$$

Discussion

The result of 3.44 kg, or about 7.6 lbs, seems about right, based on experience. You might expect to use about a 4 kg (7–10 lb) bag of ice per day. A little extra ice is required if you add any warm food or beverages.

Inspecting the conductivities in Table 14.3 shows that Styrofoam is a very poor conductor and thus a good insulator. Other good insulators include fiberglass, wool, and goose-down feathers. Like Styrofoam, these all incorporate many small pockets of air, taking advantage of air's poor thermal conductivity.

Substance	Thermal conductivity k (J/s·m·ºC)
Silver	420
Copper	390
Gold	318
Aluminum	220
Steel iron	80
Steel (stainless)	14
Ice	2.2
Glass (average)	0.84
Concrete brick	0.84
Water	0.6
Fatty tissue (without blood)	0.2
Asbestos	0.16
Plasterboard	0.16
Wood	0.08–0.16
Snow (dry)	0.10
Cork	0.042
Glass wool	0.042
Wool	0.04
Down feathers	0.025

Substance	Thermal conductivity k (J/s·m·ºC)
Air	0.023
Styrofoam	0.010

Table 14.3: Thermal Conductivities of Common Substances[7]

A combination of material and thickness is often manipulated to develop good insulators—the smaller the conductivity *k* and the larger the thickness *d*, the better. The ratio of d/k will thus be large for a good insulator. The ratio d/k is called the *R* factor. The rate of conductive heat transfer is inversely proportional to *R*. The larger the value of *R*, the better the insulation. *R* factors are most commonly quoted for household insulation, refrigerators, and the like—unfortunately, it is still in non-metric units of ft^2·°F·h/Btu, although the unit usually goes unstated (1 British thermal unit [Btu] is the amount of energy needed to change the temperature of 1.0 lb of water by 1.0 °F). A couple of representative values are an *R* factor of 11 for 3.5-in-thick fiberglass batts (pieces) of insulation and an *R* factor of 19 for 6.5-in-thick fiberglass batts. Walls are usually insulated with 3.5-in batts, while ceilings are usually insulated with 6.5-in batts. In cold climates, thicker batts may be used in ceilings and walls.

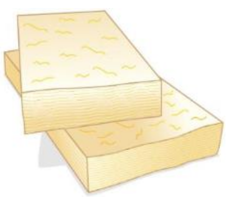


Figure 14.11: The fiberglass batt is used for insulation of walls and ceilings to prevent heat transfer between the inside of the building and the outside environment.

Note that in Table 14.3, the best thermal conductors—silver, copper, gold, and aluminum—are also the best electrical conductors, again related to the density of free electrons in them. Cooking utensils are typically made from good conductors.

Conduction is caused by the random motion of atoms and molecules. As such, it is an ineffective mechanism for heat transport over macroscopic distances and short time distances. Take, for example, the temperature on the Earth, which would be unbearably cold during the night and extremely hot during the day if heat transport in the atmosphere was to be only through conduction. In another example, car engines would overheat unless there was a more efficient way to remove excess heat from the pistons.

Convection is driven by large-scale flow of matter. In the case of Earth, the atmospheric circulation is caused by the flow of hot air from the tropics to the poles, and the flow of cold air from the poles toward the tropics. (Note that Earth's rotation causes the observed easterly flow of air in the northern hemisphere). Car engines are kept cool by the flow of water in the cooling system, with the water pump maintaining a flow of cool water to the pistons. The circulatory system is used the body: when the body overheats, the blood vessels in the skin expand (dilate), which increases the blood flow to the skin where it can be cooled by sweating. These vessels become smaller when it is cold outside and larger when it is hot (so more fluid flows, and more energy is transferred).

The body also loses a significant fraction of its heat through the breathing process.

While convection is usually more complicated than conduction, we can describe convection and do some straightforward, realistic calculations of its effects. Natural convection is driven by buoyant forces: hot air rises because density decreases as temperature increases. The house in Figure 14.12 is kept warm in this manner, as is the pot of water on the stove in Figure 14.13. Ocean currents and large-scale atmospheric circulation transfer energy from one part of the globe to another. Both are examples of natural convection.

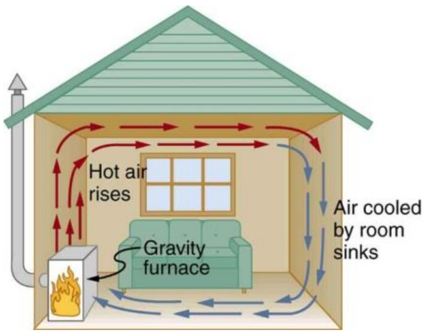


Figure 14.12: Air heated by the so-called gravity furnace expands and rises, forming a convective loop that transfers energy to other parts of the room. As the air is cooled at the ceiling and outside walls, it contracts, eventually becoming denser than room air and sinking to the floor. A properly designed heating system using natural convection, like this one, can be quite efficient in uniformly heating a home.

Figure 14.13: Convection plays an important role in heat transfer inside this pot of water. Once conducted to the inside, heat transfer to other parts of the pot is mostly by convection. The hotter water expands, decreases in density, and rises to transfer heat to other regions of the water, while colder water sinks to the bottom. This process keeps repeating.

Example - Calculating Heat Transfer by Convection: Convection of Air Through the Walls of a House

Most houses are not airtight: air goes in and out around doors and windows, through cracks and crevices, following wiring to switches and outlets, and so on. The air in a typical house is completely replaced in less than an hour. Suppose that a moderately-sized house has inside dimensions 12.0m×18.0m×3.00m high, and that all air is replaced in 30.0 min. Calculate the heat transfer per unit time in watts needed to warm the incoming cold air by 10.0ºC, thus replacing the heat transferred by convection alone.

Strategy

Heat is used to raise the temperature of air so that $Q = mc\Delta T$. The rate of heat transfer is then Q/t, where t is the time for air turnover. We are given that ΔT is 10.0ºC, but we must still find values for the mass of air and its specific heat before we can calculate Q. The specific heat of air is a weighted average of the specific heats of nitrogen and oxygen, which gives $c = c\text{p} \cong$ 1000 J/kg·ºC from Table 14.1 (note that the specific heat at constant pressure must be used for this process).

Solution

1. Determine the mass of air from its density and the given volume of the house. The density is given from the density ρ and the volume

$$m = \rho V = (1.29\text{kg/m}^3)(12.0\text{m}\times18.0\text{m}\times3.00\text{m}) = 836\text{kg}.$$

2. Calculate the heat transferred from the change in air temperature: $Q = mc\Delta T$ so that

$$Q = (836\text{kg})(1000\ \text{J/kg}\cdot{}^{\circ}\text{C})(10.0{}^{\circ}\text{C}) = 8.36\times10^6\text{J}.$$

3. Calculate the heat transfer from the heat Q and the turnover time t. Since air is turned over in t = 0.500h = 1800s, the heat transferred per unit time is

$$Q / t = (8.36\times10^6\ \text{J}) / (1800\ \text{s}) = 4.64\text{kW}$$

Discussion

This rate of heat transfer is equal to the power consumed by about forty-six 100-W light bulbs. Newly constructed homes are designed for a turnover time of 2 hours or more, rather than 30 minutes for the house of this example. Weather stripping, caulking, and improved window seals are commonly employed. More extreme measures are sometimes taken in very cold (or hot) climates to achieve a tight standard of more than 6 hours for one air turnover. Still longer turnover times are unhealthy, because a minimum amount of fresh air is necessary to supply oxygen for breathing and to dilute household pollutants. The term used for the process by which outside air leaks into the house from cracks around windows, doors, and the foundation is called “air infiltration.”

A cold wind is much more chilling than still cold air, because convection combines with conduction in the body to increase the rate at which energy is transferred away from the body. The table below gives approximate wind−chill factors, which are the temperatures of still air that produce the same rate of cooling as air of a given temperature and speed. Wind−chill factors are a dramatic reminder of convection's ability to transfer heat faster than conduction. For example, a 15.0 m/s wind at 0ºC has the chilling equivalent of still air at about −18ºC.

Moving air temperature (ºC)	**Wind speed (m/s)**				
	2	**5**	**10**	**15**	**20**
5	3	−1	−8	−10	−12
2	0	−7	−12	−16	−18
0	−2	−9	−15	−18	−20
−5	−7	−15	−22	−26	−29
−10	−12	−21	−29	−34	−36
−20	−23	−34	−44	−50	−52
−40	−44	−59	−73	−82	−84

Table 14.4: Wind-Chill Factors

Although air can transfer heat rapidly by convection, it is a poor conductor and thus a good insulator. The amount of available space for airflow determines whether air acts as an insulator or conductor. The space between the inside and outside walls of a house, for example, is about 9 cm (3.5 in) —large enough for convection to work effectively. The addition of wall insulation prevents airflow, so heat loss (or gain) is decreased. Similarly, the gap between the two panes of a double-paned window is about 1 cm, which prevents convection and takes advantage of air's low conductivity to prevent greater loss. Fur, fiber, and fiberglass also take advantage of the low conductivity of air by trapping it in spaces too small to support convection, as shown in the figure. Fur and feathers are lightweight and thus ideal for the protection of animals.

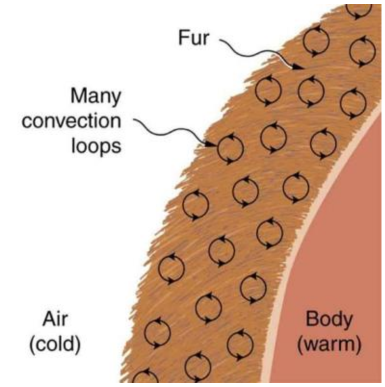


Figure 14.14: Fur is filled with air, breaking it up into many small pockets. Convection is very slow here, because the loops are so small. The low conductivity of air makes fur a very good lightweight insulator.

Some interesting phenomena happen *when convection is accompanied by a phase change.* It allows us to cool off by sweating, even if the temperature of the surrounding air exceeds body temperature. Heat from the skin is required for sweat to evaporate from the skin, but without air flow, the air becomes saturated and evaporation stops. Air flow caused by convection replaces the saturated air by dry air and evaporation continues.

Example - Calculate the Flow of Mass during Convection: Sweat-Heat Transfer away from the Body

The average person produces heat at the rate of about 120 W when at rest. At what rate must water evaporate from the body to get rid of all this energy? (This evaporation might occur when a person is sitting in the shade and surrounding temperatures are the same as skin temperature, eliminating heat transfer by other methods.)

Strategy

Energy is needed for a phase change ($Q = mL_v$). Thus, the energy loss per unit time is

$$Q/t = mL_v t = 120 \text{ W} = 120 \text{ J/s}.$$

We divide both sides of the equation by L_v to find that the mass evaporated per unit time is

$$m / t = (120 \text{ J/s}) / Lv$$

Solution

(1) Insert the value of the latent heat from Table 14.2, L_v = 2430 kJ/kg = 2430 J/g. This yields,

$$m / t = (120 \text{ J/s}) / (2430 \text{ J/g}) = 0.0494 \text{ g/s} = 2.96 \text{ g/min}.$$

Discussion

Evaporating about 3 g/min seems reasonable. This would be about 180 g (about 7 oz) per hour. If the air is very dry, the sweat may evaporate without even being noticed. A significant amount of evaporation also takes place in the lungs and breathing passages.

Another important example of the combination of phase change and convection occurs when water evaporates from the oceans. Heat is removed from the ocean when water evaporates. If the water vapor condenses in liquid droplets as clouds form, heat is released in the atmosphere. Thus, there is an overall transfer of heat from the ocean to the atmosphere. This process is the driving power behind thunderheads, those great cumulus clouds that rise as much as 20.0 km into the stratosphere. Water vapor carried in by convection condenses, releasing tremendous amounts of energy. This energy causes the air to expand and rise, where it is colder. More condensation occurs in these colder regions, which in turn drives the cloud even higher. Such a mechanism is called positive feedback, since the process reinforces and accelerates itself. These systems sometimes produce violent storms, with lightning and hail, and constitute the mechanism driving hurricanes.

The movement of icebergs is another example of convection accompanied by a phase change. Suppose an iceberg drifts from Greenland into warmer Atlantic waters. Heat is removed from the warm ocean water when the ice melts and heat is released to the land mass when the iceberg forms on Greenland.

You can feel the heat transfer from a fire and from the Sun. Similarly, you can sometimes tell that the oven is hot without touching its door or looking inside—it may just warm you as you walk by. The space between the Earth and the Sun is largely empty, without any possibility of heat transfer by convection or conduction. In these examples, heat is transferred by radiation. That is, the hot body emits electromagnetic waves that are absorbed by our skin: no medium is required for electromagnetic waves to propagate. Different names are used for electromagnetic waves of different wavelengths: radio waves, microwaves, infrared radiation, visible light, ultraviolet radiation, X-rays, and gamma rays.

The energy of electromagnetic radiation depends on the wavelength (color) and varies over a wide range: a smaller wavelength (or higher frequency) corresponds to a higher energy. Because more heat is radiated at higher temperatures, a temperature change is accompanied by a color change. Take, for example, an electrical element on a stove, which glows from red to orange, while the higher-temperature steel in a blast furnace glows from yellow to white. The radiation you feel is mostly infrared, which corresponds to a lower temperature than that of the electrical element and the steel. The radiated energy depends on its intensity, which is represented in the figure below by the height of the distribution.

All objects absorb and emit electromagnetic radiation. The rate of heat transfer by radiation is largely determined by the color of the object. Black is the most effective, and white is the least effective. People living in hot climates generally avoid wearing black clothing, for instance. Similarly, black asphalt in a parking lot will be hotter than adjacent gray sidewalk on a summer day, because black absorbs better than gray. The reverse is also true — black radiates better than gray. Thus, on a clear summer night, the asphalt will be colder than the gray sidewalk, because black radiates the energy more rapidly than gray. An *ideal radiator* is the same color as an *ideal absorber*, and captures all the radiation that falls on it. In contrast, white is a poor absorber and is also a poor radiator. A white object reflects all radiation, like a mirror. (A perfect, polished white surface is mirror-like in appearance, and a crushed mirror looks white.)

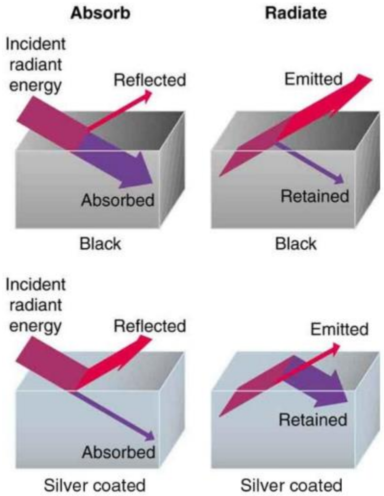


Figure 14.15: A black object is a good absorber and a good radiator, while a white (or silver) object is a poor absorber and a poor radiator. It is as if radiation from the inside is reflected back into the silver object, whereas radiation from the inside of the black object is "absorbed" when it hits the surface and finds itself on the outside and is strongly emitted.

The rate of heat transfer by emitted radiation is determined by the Stefan-Boltzmann law of radiation:

$$Q / t = \sigma e A T^4$$

where $\sigma = 5.67 \times 10^{-8} \text{J/s} \cdot \text{m}^2 \cdot \text{K}^4$ is the Stefan-Boltzmann constant, A is the surface area of the object, and T is its absolute temperature in kelvin. The symbol e stands for the emissivity of the object, which is a measure of how well it radiates. An ideal jet-black (or black body) radiator has $e = 1$, whereas a perfect reflector has $e = 0$. Real objects fall between these two values. Take, for example, tungsten light bulb filaments which have an e of about 0.5, and carbon black (a material used in printer toner), which has the (greatest known) emissivity of about 0.99.

The radiation rate is directly proportional to the *fourth power* of the absolute temperature—a remarkably strong temperature dependence. Furthermore, the radiated heat is proportional to the surface area of the object. If you knock apart the coals of a fire, there is a noticeable increase in radiation due to an increase in radiating surface area.


Figure 14.16: A thermograph of part of a building shows temperature variations, indicating where heat transfer to the outside is most severe. Windows are a major region of heat transfer to the outside of homes. (credit: U.S. Army)

Skin is a remarkably good absorber and emitter of infrared radiation, having an emissivity of 0.97 in the infrared spectrum. Thus, we are all nearly (jet) black in the infrared, in spite of the obvious variations in skin color. This high infrared emissivity is why we can so easily feel radiation on our skin. It is also the basis for the use of night scopes used by law enforcement and the military to detect human beings. Even small temperature variations can be detected because of the T^4 dependence. Images, called *thermographs*, can be used medically to detect regions of abnormally high temperature in the body, perhaps indicative of disease. Similar techniques can be used to detect heat leaks in homes Figure 14.16, optimize performance of blast furnaces, improve comfort levels in work environments, and even remotely map the Earth's temperature profile.

All objects emit and absorb radiation. The *net* rate of heat transfer by radiation (absorption minus emission) is related to both the temperature of the object and the temperature of its surroundings. Assuming that an object with a temperature T_1 is surrounded by an environment with uniform temperature T_2, the net rate of heat transfer by radiation is,

$$Q\text{net} / t = \sigma e A(T^4{}_2 - T^4{}_1)$$

where e is the emissivity of the object alone. In other words, it does not matter whether the surroundings are white, gray, or black; the balance of radiation into and out of the object depends on how well it emits and absorbs radiation. When $T_2 > T_1$, the quantity Q_{net}/t is positive; that is, the net heat transfer is from hot to cold.

Example - Calculate the Net Heat Transfer of a Person: Heat Transfer by Radiation

What is the rate of heat transfer by radiation, with an unclothed person standing in a dark room whose ambient temperature is 22.0ºC. The person has a normal skin temperature of 33.0ºC and a surface area of 1.50m^2. The emissivity of skin is 0.97 in the infrared, where the radiation takes place.

Strategy

We can solve this by using the equation for the rate of radiative heat transfer.

Solution

Insert the temperatures values T_2=295 K and T_1=306 K, so that

$$Q / t = \sigma e A(T^4_2 - T^4_1)$$

$$= (5.67\times10^{-8}\ \mathrm{J/s\cdot m^2\cdot K^4})(0.97)(1.50\ \mathrm{m^2})[(295\ \mathrm{K})^4 - (306\ \mathrm{K})^4]$$

$$= -99\ \mathrm{J/s} = -99\ \mathrm{W}.$$

Discussion

This value is a significant rate of heat transfer to the environment (note the minus sign), considering that a person at rest may produce energy at the rate of 125 W and that conduction and convection will also be transferring energy to the environment. Indeed, we would probably expect this person to feel cold. Clothing significantly reduces heat transfer to the environment by many methods, because clothing slows down both conduction and convection, and has a lower emissivity (especially if it is white) than skin.

The Earth receives almost all its energy from radiation of the Sun and reflects some of it back into outer space. Because the Sun is hotter than the Earth, the net energy flux is from the Sun to the Earth. However, the rate of energy transfer is less than the equation for the radiative heat transfer would predict because the Sun does not fill the sky. The average emissivity (*e*) of the Earth is about 0.65, but the calculation of this value is complicated by the fact that the highly reflective cloud coverage varies greatly from day to day. There is negative feedback (one in which a change produces an effect that opposes that change) between clouds and heat transfer; greater temperatures evaporate more water to form more clouds, which reflect more radiation back into space, reducing the temperature. The often-mentioned greenhouse effect is directly related to the variation of the Earth's emissivity with radiation type (see the figure given below). The greenhouse effect is a natural phenomenon responsible for providing temperatures suitable for life on Earth. The Earth's relatively constant temperature is a result of the energy balance between the incoming solar radiation and the energy radiated from the Earth. Most of the infrared radiation emitted from the Earth is absorbed by carbon dioxide (CO_2) and water (H_2O) in the atmosphere and then re-radiated back to the Earth or into outer space. Re-radiation back to the Earth maintains its surface temperature about 40ºC higher than it would be if there was no atmosphere, similar to the way glass increases temperatures in a greenhouse.

The greenhouse effect and its causes were first predicted by Eunice Newton Foote after she designed and conducted experiments on heating of different gases. After filling flasks with carbon dioxide, hydrogen, and regular air, then also modifying moisture, she placed them in the sun and carefully measured their heating and, especially, their heat retention. She discovered that the CO_2 flask gained the most temperature and held it the longest. After subsequent research, her paper "Circumstances affecting the Heat of the Sun's Rays" included conclusions that an atmosphere consisting of more carbon dioxide would be hotter resulting from the gas trapping radiation.

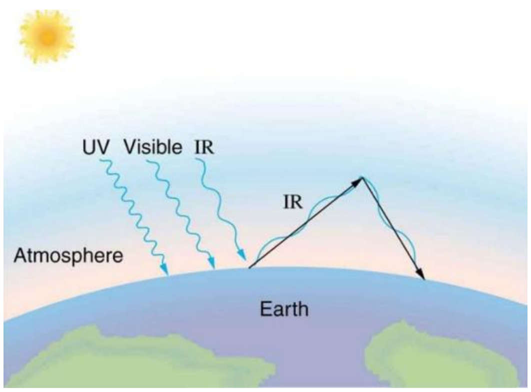


Figure 14.17: The greenhouse effect is a name given to the trapping of energy in the Earth's atmosphere by a process similar to that used in greenhouses. The atmosphere, like window glass, is transparent to incoming visible radiation and most of the Sun's infrared. These wavelengths are absorbed by the Earth and re-emitted as infrared. Since Earth's temperature is much lower than that of the Sun, the infrared radiated by the Earth has a much longer wavelength. The atmosphere, like glass, traps these longer infrared rays, keeping the Earth warmer than it would otherwise be. The amount of trapping depends on concentrations of trace gases like carbon dioxide, and a change in the concentration of these gases is believed to affect the Earth's surface temperature.

The greenhouse effect is also central to the discussion of global warming due to emission of carbon dioxide and methane (and other so-called greenhouse gases) into the Earth's atmosphere from industrial production and farming. Changes in global climate could lead to more intense storms, precipitation changes (affecting agriculture), reduction in rain forest biodiversity, and rising sea levels.

Heating and cooling are often significant contributors to energy use in individual homes. Mária Telkes, a Hungarian-born American scientist, was among the foremost developers of solar energy applications in industrial and community use. After inventing a widely deployed solar seawater distiller used on World War II life rafts, she partnered with architect Eleanor Raymond to design the first modern home to be completely heated by solar power. Air warmed on rooftop collectors transferred heat to salts, which stored the heat for later use. Telkes later worked with the Department of Energy to develop the first solar-electrically powered home. Current research efforts into developing environmentally friendly homes quite often focus on reducing conventional heating and cooling through better building materials, strategically positioning windows to optimize radiation gain from the Sun, and opening spaces to allow convection. It is possible to build a zero-energy house that allows for comfortable living in most parts of the United States with hot and humid summers and cold winters.

Conversely, dark space is very cold, about 3K (−454ºF), so that the Earth radiates energy into the dark sky. Owing to the fact that clouds have lower emissivity than either oceans or land masses, they reflect some of the radiation back to the surface, greatly reducing heat transfer into dark space, just as they greatly reduce heat transfer into the atmosphere during the day. The rate of heat transfer from soil and grasses can be so rapid that frost may occur on clear summer evenings, even in warm latitudes.

For examples and answers, please refer to OpenStax.com questions and answers given on their website or to the College Physics 2e - https://openstax.org/details/books/college-physics-2e.

Chapter 15

15.0 Objectives

At the end of this lesson, students should be able to,

- Apply first and second laws of thermodynamics.
- Interpret entropy and the second law of thermodynamics statistically.
- Identify Carnot's engine and apply knowledge to solve problems.

15.1 Introduction

This chapter covers the first and second of law thermodynamics, entropy, Carnot's engine and related calculations.

Heat transfer is energy in transit, and it can be used to do work. It can also be converted to any other form of energy. A car engine, for example, burns fuel for heat transfer into a gas. Work is done by the gas as it exerts a force through a distance, converting its energy into a variety of other forms—into the car's kinetic or gravitational potential energy; into electrical energy to run the spark plugs, radio, and lights; and back into stored energy in the car's battery. But most of the heat transfer produced from burning fuel in the engine does not do work on the gas. Rather, the energy is released into the environment, implying that the engine is quite inefficient.

It is often said that modern gasoline engines cannot be made to be significantly more efficient. We hear the same about heat transfer to electrical energy in large power stations, whether they are coal, oil, natural gas, or nuclear powered. Why is that the case? Is the inefficiency caused by design problems that could be solved with better engineering and superior materials? Is it part of some money-making conspiracy by those who sell energy? Actually, the truth is more interesting, and reveals much about the nature of heat transfer.

Basic physical laws govern how heat transfer for doing work takes place and place insurmountable limits on its efficiency. This chapter will explore these laws as well as many applications and concepts associated with them. These topics are part of *thermodynamics*—the study of heat transfer and its relationship to doing work.

If we are interested in how heat transfer is converted into doing work, then the conservation of energy principle is important. The first law of thermodynamics applies the conservation of energy principle to systems where heat transfer and doing work are the methods of transferring energy into and out of the system. The first law of thermodynamics states that the change in internal energy of a system equals the net heat transfer *into* the system minus the net work done *by* the system. In equation form, the first law of thermodynamics is

$$\Delta U = Q - W$$

Here ΔU is the *change in internal energy* U of the system. Q is the *net heat transferred into the system*—that is, Q is the sum of all heat transfer into and out of the system. W is the *net work done by the system*—that is, W is the sum of all work done on or by the system. We use the following sign conventions: if Q is positive, then there is a net heat transfer into the system; if W is positive, then there is net work done by the system. So positive Q adds energy to the system and positive W takes energy from the system. Thus $\Delta U = Q - W$. Note also that if more heat transfer into the system occurs than work done, the difference is stored as internal energy. Heat engines are a good example of this—heat transfer into them takes place so that they can do work. (See Figure 15.1) We will now examine Q, W, and ΔU further.

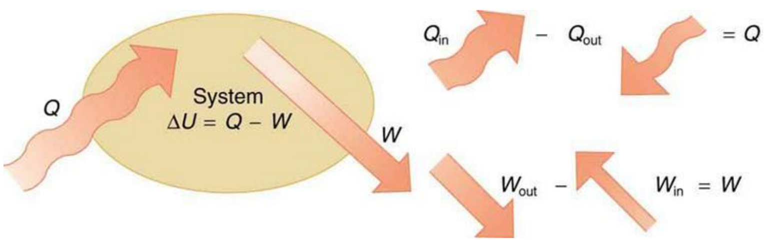


Figure 15.1: The first law of thermodynamics is the conservation-of-energy principle stated for a system where heat and work are the methods of transferring energy for a system in thermal equilibrium. Q represents the net heat transfer—it is the sum of all heat transfers into and out of the system. Q is positive for net heat transfer *into* the system. W is the total work done on and by the system. W is positive when more work is done *by* the system than on it. The change in the internal energy of the system, ΔU, is related to heat and work by the first law of thermodynamics, $\Delta U = Q - W$.

15.2 First Law of Thermodynamics and the Law of Conservation of Energy

The first law of thermodynamics is actually the law of conservation of energy stated in a form most useful in thermodynamics. The first law gives the relationship between heat transfer, work done, and the change in internal energy of a system.

Heat *Q* and Work *W*

Heat transfer (Q) and doing work (W) are the two everyday means of bringing energy into or taking energy out of a system. The processes are quite different. Heat transfer, a less organized process, is driven by temperature differences. Work, a quite organized process, involves a macroscopic force exerted through a distance. Nevertheless, heat and work can produce identical results. For example, both can cause a temperature increase. Heat transfer into a system, such as when the Sun warms the air in a bicycle tire, can increase its temperature, and so can work done on the system, as when the bicyclist pumps air into the tire. Once the temperature increase has occurred, it is impossible to tell whether it was caused by heat transfer or by doing work. This uncertainty is an important point. Heat transfer and work are both energy in transit—neither is stored as such in a system. However, both can change the internal energy U of a system. Internal energy is a form of energy completely different from either heat or work.

Internal Energy

We can think about the internal energy of a system in two different but consistent ways. The first is the atomic and molecular view, which examines the system on the atomic and molecular scale. The internal energy U of a system is the sum of the kinetic and potential energies of its atoms and molecules. Recall that kinetic plus potential energy is called mechanical energy. Thus, internal energy is the sum of atomic and molecular mechanical energy. Because it is impossible to keep track of all individual atoms and molecules, we must deal with averages and distributions. A second way to view the internal energy of a system is in terms of its macroscopic characteristics, which are very similar to atomic and molecular average values.

Macroscopically, we define the change in internal energy ΔU to be that given by the first law of thermodynamics:

$$\Delta U = Q - W.$$

Many detailed experiments have verified that $\Delta U = Q - W$, where ΔU is the change in total kinetic and potential energy of all atoms and molecules in a system. It has also been determined experimentally that the internal energy U of a system depends only on the state of the system and *not how it reached that state*. More specifically, U is found to be a function of a few macroscopic quantities (pressure, volume, and temperature, for example), independent of past history such as whether there has been heat transfer or work done. This independence means that if we know the state of a system, we can calculate changes in its internal energy U from a few macroscopic variables.

In thermodynamics, we often use the macroscopic picture when making calculations of how a system behaves, while the atomic and molecular picture gives underlying explanations in terms of averages and distributions. For example, in the topic of entropy, calculations will be made using the atomic and molecular view.

To get a better idea of how to think about the internal energy of a system, let us examine a system going from State 1 to State 2. The system has internal energy U_1 in State 1, and it has internal energy U_2 in State 2, no matter how it got to either state. So the change in internal energy $\Delta U = U_2 - U_1$ is independent of what caused the change. In other words, ΔU *is independent of path*. By path, we mean the method of getting from the starting point to the ending point. Why is this independence important? Note that $\Delta U = Q - W$. Both Q and W *depend on path*, but ΔU does not. This path independence means that internal energy U is easier to consider than either heat transfer or work done.

Example - Calculating Change in Internal Energy: The Same Change in *U* is Produced by Two Different Processes

(a) Suppose there is heat transfer of 40.00 J to a system, while the system does 10.00 J of work. Later, there is heat transfer of 25.00 J out of the system while 4.00 J of work is done on the system. What is the net change in internal energy of the system?

(b) What is the change in internal energy of a system when a total of 150.00 J of heat transfer occurs out of (from) the system and 159.00 J of work is done on the system? (See Figure 15.2).

Strategy

In part (a), we must first find the net heat transfer and net work done from the given information. Then the first law of thermodynamics ($\Delta U = Q - W$) can be used to find the change in internal energy. In part (b), the net heat transfer and work done are given, so the equation can be used directly.

Solution for (a)

The net heat transfer is the heat transfer into the system minus the heat transfer out of the system, or

$$Q = 40.00\text{J} - 25.00\text{ J} = 15.00\text{J}$$

Similarly, the total work is the work done by the system minus the work done on the system, or

$$W = 10.00\text{J} - 4.00\text{J} = 6.00\text{ J}$$

Thus, the change in internal energy is given by the first law of thermodynamics:

$$\Delta U = Q - W = 15.00\text{J} - 6.00\text{J} = 9.00\text{ J}$$

We can also find the change in internal energy for each of the two steps. First, consider 40.00 J of heat transfer in and 10.00 J of work out, or

$$\Delta U_1 = Q_1 - W_1 = 40.00\text{J} - 10.00\text{J} = 30.00\text{ J}$$

Now consider 25.00 J of heat transfer out and 4.00 J of work in, or

The total change is the sum of these two steps, or

$$\Delta U = \Delta U_1 + \Delta U_2 = 30.00\text{J} + (-21.00\text{J}) = 9.00\text{ J}$$

Discussion on (a)

No matter whether you look at the overall process or break it into steps, the change in internal energy is the same.

Solution for (b)

Here the net heat transfer and total work are given directly to be $Q = -150.00$ J and $W = -159.00$ J, so that

$$\Delta U = Q - W = -150.00\text{ J} - (-159.00\text{ J}) = 9.00\text{ J}.$$

Discussion on (b)

A very different process in part (b) produces the same 9.00-J change in internal energy as in part (a). Note that the change in the system in both parts is related to ΔU and not to the individual *Q*s or *W*s involved. The system ends up in the *same* state in both (a) and (b). Parts (a) and (b) present two different paths for the system to follow between the same starting and ending points, and the change in internal energy for each is the same—it is independent of path.

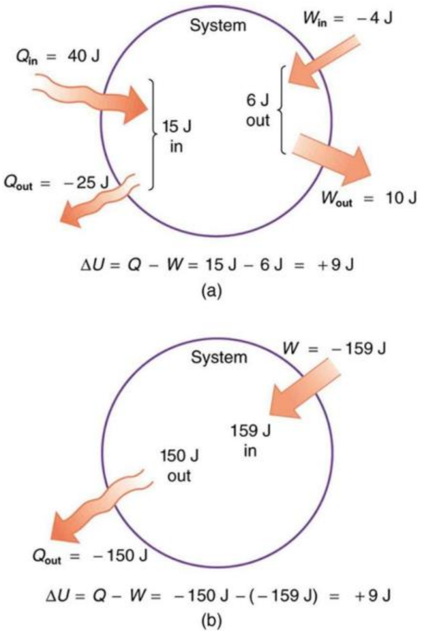


Figure 15.2: Two different processes produce the same change in a system. (a) A total of 15.00 J of heat transfer occurs into the system, while work takes out a total of 6.00 J. The change in internal energy is $\Delta U = Q - W = 9.00$ J. (b) Heat transfer removes 150.00 J from the system while work puts 159.00 J into it, producing an increase of 9.00 J in internal energy. If the system starts out in the same state in (a) and (b), it will end up in the same final state in either case—its final state is related to internal energy, not how that energy was acquired.

*** In this equation, $\Delta U = Q - W$ – *the system is doing work on the surrounding is taken as positive, hence this equation appears as Q – W. This has been the framework of reference throughout this chapter.*

**** If the reference point is for the system, then work done on the system is positive. Then the equation is $\Delta U = Q + W$. You will notice this mostly in chemistry.*

**** Nevertheless, in both cases, the frame of references has to be consistent when doing the calculations.*

15.3 Human Metabolism and the First Law of Thermodynamics

Human metabolism is the conversion of food into heat transfer, work, and stored fat. Metabolism is an interesting example of the first law of thermodynamics in action. We now take another look at these topics via the first law of thermodynamics. Considering the body as the system of interest, we can use the first law to examine heat transfer, doing work, and internal energy in activities ranging from sleep to heavy exercise. What are some of the major characteristics of heat transfer, doing work, and energy in the body? For one, body temperature is normally kept constant by heat transfer to the surroundings. This means Q is negative. Another fact is that the body usually does work on the outside world. This means W is positive. In such situations, then, the body loses internal energy, since $\Delta U=Q-W$ is negative.

Now consider the effects of eating. Eating increases the internal energy of the body by adding chemical potential energy (this is an unromantic view of a good steak). The body *metabolizes* all the food we consume. Basically, metabolism is an oxidation process in which the chemical potential energy of food is released. This implies that food input is in the form of work. Food energy is reported in a special unit, known as the Calorie. This energy is measured by burning food in a calorimeter, which is how the units are determined.

In chemistry and biochemistry, one calorie (spelled with a *lowercase* c) is defined as the energy (or heat transfer) required to raise the temperature of one gram of pure water by one degree Celsius. Nutritionists and weightwatchers tend to use the *dietary* calorie, which is frequently called a Calorie (spelled with a *capital* C). One food Calorie is the energy needed to raise the temperature of one *kilogram* of water by one degree Celsius. This means that one dietary Calorie is equal to one kilocalorie for the chemist, and one must be careful to avoid confusion between the two.

Again, consider the internal energy the body has lost. There are three places this internal energy can go—to heat transfer, to doing work, and to stored fat (a tiny fraction also goes to cell repair and growth). Heat transfer and doing work take internal energy out of the body, and food puts it back. If you eat just the right amount of food, then your average internal energy remains constant. Whatever you lose to heat transfer and doing work is replaced by food, so that, in the long run, $\Delta U=0$. If you overeat repeatedly, then ΔU is always positive, and your body stores this extra internal energy as fat. The reverse is true if you eat too little. If ΔU is negative for a few days, then the body metabolizes its own fat to maintain body temperature and do work that takes energy from the body. This process is how dieting produces weight loss.

Life is not always this simple, as any dieter knows. The body stores fat or metabolizes it only if energy intake changes for a period of several days. Once you have been on a major diet, the next one is less successful because your body alters the way it responds to low energy intake. Your basal metabolic rate (BMR) is the rate at which food is converted into heat transfer and work done while the body is at complete rest. The body adjusts its basal metabolic rate to partially compensate for over-eating or under-eating. The body will decrease the metabolic rate rather than eliminate its own fat to replace lost food intake. You will chill more easily and feel less energetic as a result of the lower metabolic rate, and you will not lose weight as fast as before. Exercise helps to lose weight, because it produces both heat transfer from your body and work, and raises your metabolic rate even when you are at rest. Weight loss is also aided by the quite low efficiency of the body in converting internal energy to work, so that the loss of internal energy resulting from doing work is much greater than the work done. It should be noted, however, that living systems are not in thermal equilibrium.

The body provides us with an excellent indication that many thermodynamic processes are *irreversible*. An irreversible process can go in one direction but not the reverse, under a given set of conditions. For example, although body fat can be converted to do work and produce heat transfer, work done on the body and heat transfer into it cannot be converted to body fat. Otherwise, we could skip lunch by sunning ourselves or by walking downstairs. Another example of an irreversible thermodynamic process is photosynthesis. This process is the intake of one form of energy—light—by plants and its conversion to chemical potential energy. Both applications of the first law of thermodynamics are illustrated in Figure 15.3. One great advantage of conservation laws such as the first law of thermodynamics is that they accurately describe the beginning and ending points of complex processes, such as metabolism and photosynthesis, without regard to the complications in between. Table 15.1 presents a summary of terms relevant to the first law of thermodynamics.

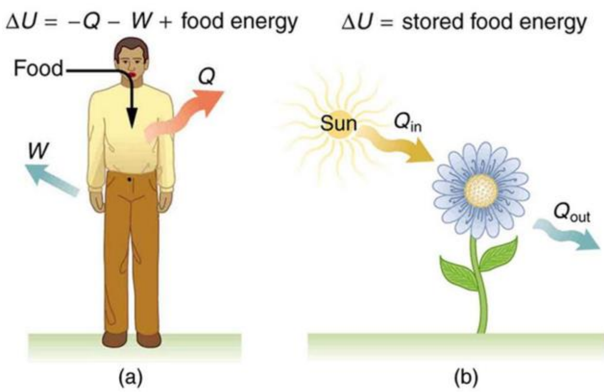


Figure 15.3: (a) The first law of thermodynamics applied to metabolism. Heat transferred out of the body (Q) and work done by the body (W) remove internal energy, while food intake replaces it. (Food intake may be considered as work done on the body.) (b) Plants convert part of the radiant heat transfer in sunlight to stored chemical energy, a process called photosynthesis.

Term	Definition
U	Internal energy—the sum of the kinetic and potential energies of a system's atoms and molecules. Can be divided into many subcategories, such as thermal and chemical energy. Depends only on the state of a system (such as its P, V, and T), not on how the energy entered the system. Change in internal energy is path independent.
Q	Heat—energy transferred because of a temperature difference. Characterized by random molecular motion. Highly dependent on path. Q entering a system is positive.
W	Work—energy transferred by a force moving through a distance. An organized, orderly process. Path dependent. W done by a system (either against an external force or to increase the volume of the system) is positive.

Table 15.1: Summary of Terms for the First Law of Thermodynamics, $\Delta U = Q - W$

One of the most important things we can do with heat transfer is to use it to do work for us. Such a device is called a heat engine. Car engines and steam turbines that generate electricity are examples of heat engines. Figure 15.4 shows schematically how the first law of thermodynamics applies to the typical heat engine.

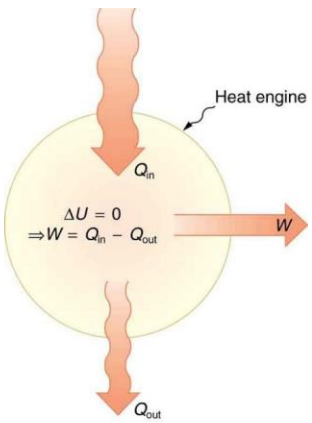


Figure 15.4: Schematic representation of a heat engine, governed, of course, by the first law of thermodynamics. It is impossible to devise a system where $Q_{out}=0$, that is, in which no heat transfer occurs to the environment.

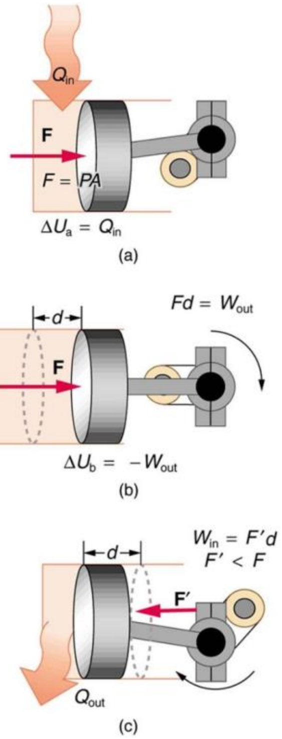


Figure 15.5: (a) Heat transfer to the gas in a cylinder increases the internal energy of the gas, creating higher pressure and temperature. (b) The force exerted on the movable cylinder does work as the gas expands. Gas pressure and temperature decrease when it expands, indicating that the gas's internal energy has been decreased by doing work. (c) Heat transfer to the environment further reduces pressure in the gas so that the piston can be more easily returned to its starting position.

The illustrations above show one of the ways in which heat transfer does work. Fuel combustion produces heat transfer to a gas in a cylinder, increasing the pressure of the gas and thereby the force it exerts on a movable piston. The gas does work on the outside world, as this force moves the piston through some distance. Heat transfer to the gas cylinder results in work being done. To repeat this process, the piston needs to be returned to its starting point. Heat transfer now occurs from the gas to the surroundings so that its pressure decreases, and a force is exerted by the surroundings to push the piston back through some distance. Variations of this process are employed daily in hundreds of millions of heat engines. We will examine heat engines in detail in the next section. In this section, we consider some of the simpler underlying processes on which heat engines are based.

15.4 PV Diagrams and their Relationship to Work Done on or by a Gas

A process by which a gas does work on a piston at constant pressure is called an isobaric process. Since the pressure is constant, the force exerted is constant and the work done is given as

$$P\Delta V$$

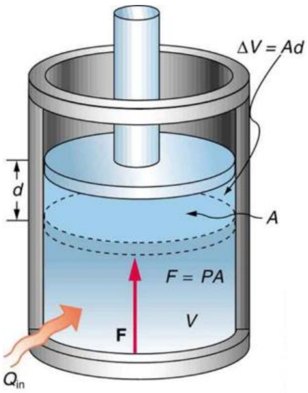


Figure 15.6: An isobaric expansion of a gas requires heat transfer to keep the pressure constant. Since pressure is constant, the work done is $P\Delta V$.

$$W = Fd$$

See the symbols as shown in Figure 15.6. Now $F = PA$, and so

$$W = PAd.$$

Because the volume of a cylinder is its cross-sectional area A times its length d, we see that $Ad = \Delta V$, the change in volume; thus,

$$W = P\Delta V \text{ (isobaric process).}$$

Note that if ΔV is positive, then W is positive, meaning that work is done *by* the gas on the outside world.

(Note that the pressure involved in this work that we've called P is the pressure of the gas *inside* the tank. If we call the pressure outside the tank Pext, an expanding gas would be working *against* the external pressure; the work done would therefore be $W = -P\text{ext}\Delta V$ (isobaric process). Many texts use this definition of work, and not the definition based on internal pressure, as the basis of the First Law of Thermodynamics. This definition reverses the sign conventions for work, and results in a statement of the first law that becomes $\Delta E_{int} = Q+W$.)

It is not surprising that $W = P\Delta V$, since we have already noted in our treatment of fluids that pressure is a type of potential energy per unit volume and that pressure in fact has units of energy divided by volume. We also noted in our discussion of the ideal gas law that *PV* has units of energy. In this case, some of the energy associated with pressure becomes work.

Figure 15.7 shows a graph of pressure versus volume (that is, a *PV* diagram for an isobaric process. You can see in the figure that the work done is the area under the graph. This property of *PV* diagrams is very useful and broadly applicable: *the work done on or by a system in going from one state to another equals the area under the curve on a PV diagram*.

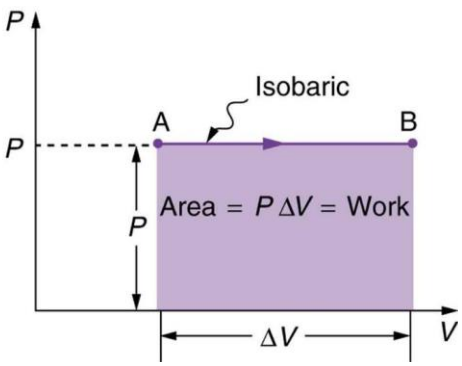


Figure 15.7: A graph of pressure versus volume for a constant-pressure, or isobaric, process, such as the one shown in Figure 15.6. The area under the curve equals the work done by the gas, since $W=P\Delta V$.

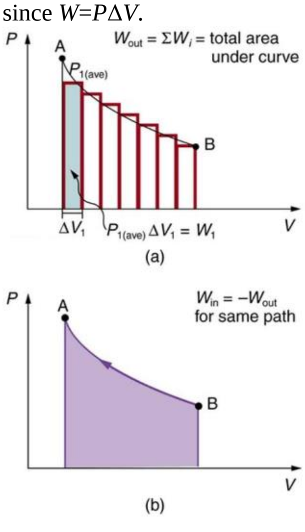


Figure 15.8: (a) A *PV* diagram in which pressure varies as well as volume. The work done for each interval is its average pressure times the change in volume, or the area under the curve over that interval. Thus the total area under the curve equals the total work done. (b) Work must be done on the system to follow the reverse path. This is interpreted as a negative area under the curve.

We can see where this leads by considering Figure 15.8 (a), which shows a more general process in which both pressure and volume change. The area under the curve is closely approximated by dividing it into strips, each having an average constant pressure P_i(ave). The work done is $W_i = P_{i(\text{ave})}\Delta V_i$ for each strip, and the total work done is the sum of the W_i. Thus the total work done is the total area under the curve. If the path is reversed, as in Figure 15.8 (b), then work is done on the system. The area under the curve in that case is negative, because ΔV is negative.

PV diagrams clearly illustrate that *the work done depends on the path taken and not just the endpoints*. This path dependence is seen in Figure 15.9 (a), where more work is done in going from A to C by the path via point B than by the path via point D. The vertical paths, where volume is constant, are called isochoric processes. Since volume is constant, $\Delta V=0$, and no work is done in an isochoric process. Now, if the system follows the cyclical path ABCDA, as in Figure 15.9 (b), then the total work done is the area inside the loop. The negative area below path CD subtracts, leaving only the area inside the rectangle. In fact, the work done in any cyclical process (one that returns to its starting point) is the area inside the loop it forms on a *PV* diagram, as Figure 15.9 (c) illustrates for a general cyclical process. Note that the loop must be traversed in the clockwise direction for work to be positive—that is, for there to be a net work output.

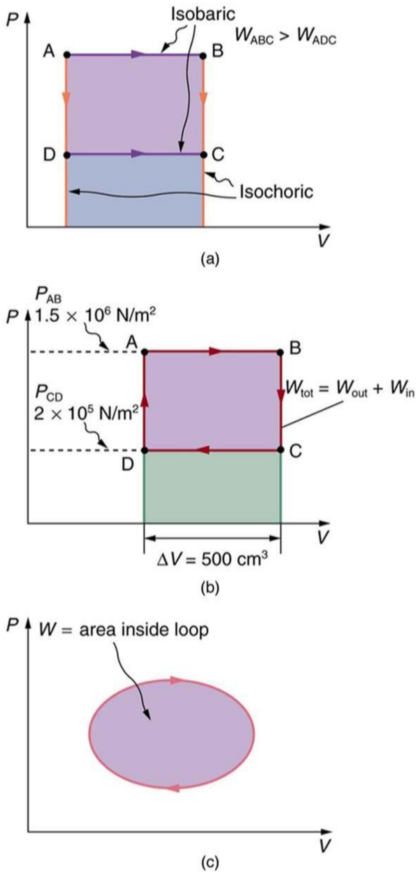


Figure 15.9: (a) The work done in going from A to C depends on path. The work is greater for the path ABC than for the path ADC, because the former is at higher pressure. In both cases, the work done is the area under the path. This area is greater for path ABC. (b) The total work done in the cyclical process ABCDA is the area inside the loop, since the negative area below CD subtracts out, leaving just the area inside the rectangle. (The values given for the pressures and the change in volume are intended for use in the example below.) (c) The area inside any closed loop is the work done in the cyclical process. If the loop is traversed in a clockwise direction, *W* is positive—it is work done on the outside environment. If the loop is traveled in a counter-clockwise direction, *W* is negative—it is work that is done to the system.

Example - Total Work Done in a Cyclical Process Equals the Area Inside the Closed Loop on a *PV* Diagram

Calculate the total work done in the cyclical process ABCDA shown in Figure 15.9 (b) by the following two methods to verify that work equals the area inside the closed loop on the *PV* diagram. (Take the data in the figure to be precise to three significant figures.) (a) Calculate the work done along each segment of the path and add these values to get the total work. (b) Calculate the area inside the rectangle ABCDA.

Strategy

To find the work along any path on a *PV* diagram, you use the fact that work is pressure times change in volume, or $W = P\Delta V$. So, in part (a), this value is calculated for each leg of the path around the closed loop.

Solution for (a)

The work along path AB is,

$$W_{AB} = P_{AB}\Delta V_{AB}$$

$$= (1.50\times10^{6}\text{N/m}^{2})\ (5.00\times10^{-4}\text{m}^{3}) = 750\text{J}$$

Since the path BC is isochoric, $\Delta V_{BC} = 0$, and so $W_{BC} = 0$. The work along path CD is negative, since ΔV_{CD} is negative (the volume decreases). The work is,

$$W_{CD} = P_{CD}\Delta V_{CD}$$

$$(2.00\times10^{5}\text{N/m}^{2})(-5.00\times10^{-4}\text{m}^{3}) = -100\text{J}$$

Again, since the path DA is isochoric, $\Delta V_{DA} = 0$, and so $W_{DA} = 0$. Now the total work is

$$W = W_{AB} + W_{BC} + W_{CD} + W_{DA}$$

$$= 750\text{ J} + 0 + (-100\text{J}) + 0 = 650\text{ J}$$

Solution for (b)

The area inside the rectangle is its height times its width, or

$$\text{area} = (P_{AB} - P_{CD})\Delta V$$

$$= [(1.50\times10^{6}\text{N/m}^{2}) - (2.00\times10^{5}\text{N/m}^{2})]\ (5.00\times10^{-4}\text{m}^{3})$$
$$= 650\text{ J}$$

Thus,

$$\text{area} = 650\text{J} = W$$

Discussion

The result, as anticipated, is that the area inside the closed loop equals the work done. The area is often easier to calculate than is the work done along each path. It is also convenient to visualize the area inside different curves on *PV* diagrams in order to see which processes might produce the most work. Recall that work can be done to the system, or by the system, depending on the sign of *W*. A positive *W* is work that is done by the system on the outside environment; a negative *W* represents work done by the environment on the system.

Figure 15.10 (a) shows two other important processes on a *PV* diagram. For comparison, both are shown starting from the same point A. The upper curve ending at point B is an isothermal process—that is, one in which temperature is kept constant. If the gas behaves like an ideal gas, as is often the case, and if no phase change occurs, then $PV = nRT$. Since T is constant, PV is a constant for an isothermal process. We ordinarily expect the temperature of a gas to decrease as it expands, and so we correctly suspect that heat transfer must occur from the surroundings to the gas to keep the temperature constant during an isothermal expansion. To show this more rigorously for the special case of a monatomic ideal gas, we note that the average kinetic energy of an atom in such a gas is given by,

$$\frac{1}{2} m v_{av}^2 = 3/2(kT)$$

The kinetic energy of the atoms in a monatomic ideal gas is its only form of internal energy, and so its total internal energy E_{int} is

$$E_{int} = N(\frac{1}{2} m v_{\mathrm{av}}^2) = 3/2(NkT) \text{ (monatomic ideal gas)},$$

where N is the number of atoms in the gas. This relationship means that the internal energy of an ideal monatomic gas is constant during an isothermal process—that is, $E_{int} = 0$. If the internal energy does not change, then the net heat transfer into the gas must equal the net work done by the gas. That is, because $\Delta E_{int} = Q - W = 0$ here, $Q = W$. We must have just enough heat transfer to replace the work done. An isothermal process is inherently slow, because heat transfer occurs continuously to keep the gas temperature constant at all times and must be allowed to spread through the gas so that there are no hot or cold regions.

Note that diatomic gases, such as those in air, have different formulas for average kinetic energy of an atom and total internal energy.

Also shown in Figure 15.10 (a) is a curve AC for an adiabatic process, defined to be one in which there is no heat transfer—that is, $Q = 0$. Processes that are nearly adiabatic can be achieved either by using very effective insulation or by performing the process so fast that there is little time for heat transfer. Temperature must decrease during an adiabatic expansion process, since work is done at the expense of internal energy:

$$E_{int} = (3/2)NkT$$

(You might have noted that a gas released into atmospheric pressure from a pressurized cylinder is substantially colder than the gas in the cylinder.) In fact, because $Q = 0$, $\Delta E_{int} = -W$ for an adiabatic process. Lower temperature results in lower pressure along the way, so that curve AC is lower than curve AB, and less work is done. If the path ABCA could be followed by cooling the gas from B to C at constant volume (isochorically), Figure 15.10 (b), there would be a net work output.

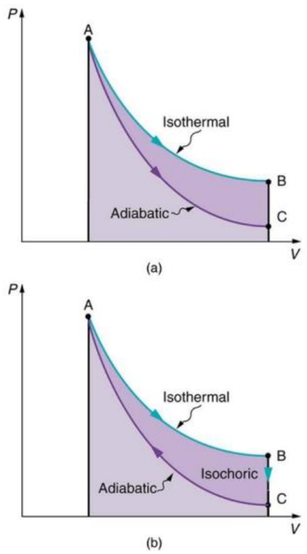


Figure 15.10: (a) The upper curve is an isothermal process (ΔT=0), whereas the lower curve is an adiabatic process (Q=0). Both start from the same point A, but the isothermal process does more work than the adiabatic because heat transfer into the gas takes place to keep its temperature constant. This keeps the pressure higher all along the isothermal path than along the adiabatic path, producing more work. The adiabatic path thus ends up with a lower pressure and temperature at point C, even though the final volume is the same as for the isothermal process. (b) The cycle ABCA produces a net work output.

15. 5 Reversible Processes

Both isothermal and adiabatic processes such as shown in Figure 15.10 are reversible in principle. A reversible process is one in which both the system and its environment can return to exactly the states they were in by following the reverse path. The reverse isothermal and adiabatic paths are BA and CA, respectively. Real macroscopic processes are never exactly reversible. In the previous examples, our system is a gas, and its environment is the piston, cylinder, and the rest of the universe. If there are any energy-dissipating mechanisms, such as friction or turbulence, then heat transfer to the environment occurs for either direction of the piston. So, for example, if the path BA is followed and there is friction, then the gas will be returned to its original state but the environment will not—it will have been heated in both directions. Reversibility requires the direction of heat transfer to reverse for the reverse path. Since dissipative mechanisms cannot be completely eliminated, real processes cannot be reversible.

There must be reasons that real macroscopic processes cannot be reversible. We can imagine them going in reverse. For example, heat transfer occurs spontaneously from hot to cold and never spontaneously the reverse. Yet it would not violate the first law of thermodynamics for this to happen. In fact, all spontaneous processes, such as bubbles bursting, never go in reverse. There is a second thermodynamic law that forbids them from going in reverse. When we study this law, we will learn something about nature and also find that such a law limits the efficiency of heat engines. We will find that heat engines with the greatest possible theoretical efficiency would have to use reversible processes, and even they cannot convert all heat transfer into doing work. Table 15.2 summarizes the simpler thermodynamic processes and their definitions.

Isobaric	Constant pressure $W=P\Delta V$
Isochoric	Constant volume $W=0$
Isothermal	Constant temperature $Q=W$
Adiabatic	No heat transfer $Q=0$

Table 15.2: Summary of Simple Thermodynamic Processes

15.6 States of Matter

The second law of thermodynamics deals with the direction taken by spontaneous processes. Many processes occur spontaneously in one direction only—that is, they are irreversible, under a given set of conditions. Although irreversibility is seen in day-to-day life—a broken glass does not resume its original state, for instance—complete irreversibility is a statistical statement that cannot be seen during the lifetime of the universe. More precisely, an irreversible process is one that depends on path. If the process can go in only one direction, then the reverse path differs fundamentally and the process cannot be reversible. For example, as noted in the previous section, heat involves the transfer of energy from higher to lower temperature. A cold object in contact with a hot one never gets colder, transferring heat to the hot object and making it hotter. Furthermore, mechanical energy, such as kinetic energy, can be completely converted to thermal energy by friction, but the reverse is impossible. A hot stationary object never spontaneously cools off and starts moving. Yet another example is the expansion of a puff of gas introduced into one corner of a vacuum chamber. The gas expands to fill the chamber, but it never regroups in the corner. The random motion of the gas molecules could take them all back to the corner, but this is never observed to happen. (See Figure 15.11)

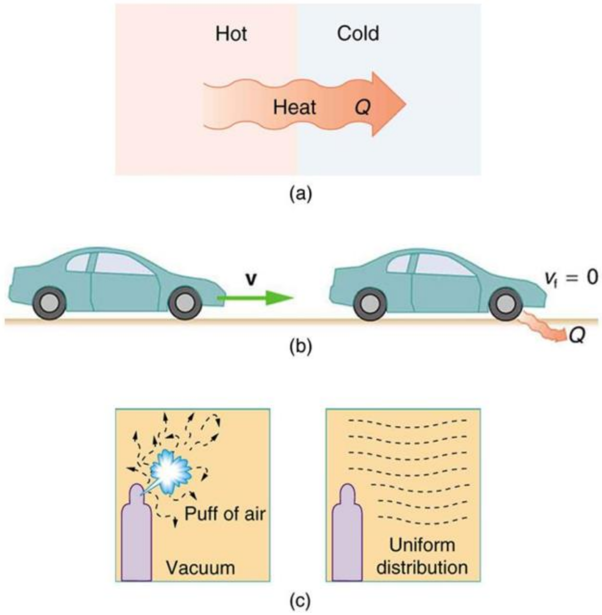


Figure 15.11: Examples of one-way processes in nature. (a) Heat transfer occurs spontaneously from hot to cold and not from cold to hot. (b) The brakes of this car convert its kinetic energy to heat transfer to the environment. The reverse process is impossible. (c) The burst of gas let into this vacuum chamber quickly expands to uniformly fill every part of the chamber. The random motions of the gas molecules will never return them to the corner.

The fact that certain processes never occur suggests that there is a law forbidding them to occur. The first law of thermodynamics would allow them to occur—none of those processes violate conservation of energy. The law that forbids these processes is called the second law of thermodynamics. We shall see that the second law can be stated in many ways that may seem different, but which in fact are equivalent. Like all natural laws, the second law of thermodynamics gives insights into nature, and its several statements imply that it is broadly applicable, fundamentally affecting many apparently disparate processes.

The already familiar direction of heat transfer from hot to cold is the basis of our first version of the second law of thermodynamics.

15.7 The Second Law of Thermodynamics (first expression)

Heat transfer occurs spontaneously from higher- to lower-temperature bodies but never spontaneously in the reverse direction.

Another way of stating this: It is impossible for any process to have as its sole result heat transfer from a cooler to a hotter object.

Heat Engines

Now let us consider a device that uses heat transfer to do work. As noted in the previous section, such a device is called a heat engine, and one is shown schematically in Figure 15.12 (b). Gasoline and diesel engines, jet engines, and steam turbines are all heat engines that do work by using part of the heat transfer from some source. Heat transfer from the hot object (or hot reservoir) is denoted as Q_h, while heat transfer into the cold object (or cold reservoir) is Q_c, and the work done by the engine is W. The temperatures of the hot and cold reservoirs are T_h and T_c, respectively.

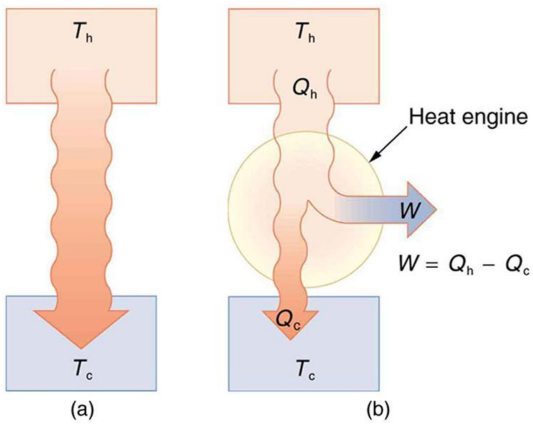


Figure 15.12: (a) Heat transfer occurs spontaneously from a hot object to a cold one, consistent with the second law of thermodynamics. (b) A heat engine, represented here by a circle, uses part of the heat transfer to do work. The hot and cold objects are called the hot and cold reservoirs. Q_h is the heat transfer out of the hot reservoir, W is the work output, and Q_c is the heat transfer into the cold reservoir.

Because the hot reservoir is heated externally, which is energy intensive, it is important that the work is done as efficiently as possible. In fact, we would like W to equal Q_h, and for there to be no heat transfer to the environment ($Q_c = 0$). Unfortunately, this is impossible. The second law of thermodynamics also states, with regard to using heat transfer to do work (the second expression of the second law):

The Second Law of Thermodynamics (second expression)

It is impossible in any system for heat transfer from a reservoir to completely convert to work in a cyclical process in which the system returns to its initial state.

A cyclical process brings a system, such as the gas in a cylinder, back to its original state at the end of every cycle. Most heat engines, such as reciprocating piston engines and rotating turbines, use cyclical processes. The second law, just stated in its second form, clearly states that such engines cannot have perfect conversion of heat transfer into work done. Before going into the underlying reasons for the limits on converting heat transfer into work, we need to explore the relationships among W, Q_h, and Q_c, and to define the efficiency of a cyclical heat engine. As noted, a cyclical process brings the system back to its original condition at the end of every cycle. Such a system's internal energy E_{int} is the same at the beginning and end of every cycle—that is, $E_{int} = 0$. The first law of thermodynamics states that

$$E_{int} = Q - W$$

where Q is the *net* heat transfer during the cycle ($Q = Q_h - Q_c$) and W is the net work done by the system. Since $E_{int} = 0$ for a complete cycle, we have

$$0 = Q - W$$

so that

$$W = Q$$

Thus the net work done by the system equals the net heat transfer into the system, or

$$W = Q_h - Q_c \text{ (cyclical process)}$$

just as shown schematically in Figure 15.12 (b). The problem is that in all processes, there is some heat transfer Q_c to the environment—and usually a very significant amount at that.

In the conversion of energy to work, we are always faced with the problem of getting less out than we put in. We define *conversion efficiency Eff* to be the ratio of useful work output to the energy input (or, in other words, the ratio of what we get to what we spend). In that spirit, we define the efficiency of a heat engine to be its net work output W divided by heat transfer to the engine Q_h; that is,

$$Eff = W/Q_h$$

Since $W = Q_h - Q_c$ in a cyclical process, we can also express this as

$$Eff = (Q_h - Q_c) / Q_h = 1 - (Q_c / Q_h) \text{ (cyclical process)},$$

making it clear that an efficiency of 1, or 100%, is possible only if there is no heat transfer to the environment ($Q_c = 0$). Note that all Qs are positive. The direction of heat transfer is indicated by a plus or minus sign. For example, Q_c is out of the system and so is preceded by a minus sign.

Example - Daily Work Done by a Coal-Fired Power Station, Its Efficiency and Carbon Dioxide Emissions

A coal-fired power station is a huge heat engine. It uses heat transfer from burning coal to do work to turn turbines, which are used to generate electricity. In a single day, a large coal power station has 2.50×10^{14}J of heat transfer from coal and 1.48×10^{14}J of heat transfer into the environment. (a) What is the work done by the power station? (b) What is the efficiency of the power station? (c) In the combustion process, the following chemical reaction occurs: $C+O_2\rightarrow CO_2$. This implies that every 12 kg of coal puts 12 kg + 16 kg + 16 kg = 44 kg of carbon dioxide into the atmosphere. Assuming that 1 kg of coal can provide 2.5×10^{6}J of heat transfer upon combustion, how much CO_2 is emitted per day by this power plant?

Strategy for (a)

We can use $W = Q_h - Q_c$ to find the work output W, assuming a cyclical process is used in the power station. In this process, water is boiled under pressure to form high-temperature steam, which is used to run steam turbine-generators, and then condensed back to water to start the cycle again.

Solution for (a)

Work output is given by:

$$W = Q_h - Q_c.$$

Substituting the given values:

$$W = 2.50\times10^{14}\text{J} - 1.48\times10^{14}\text{J}$$

$$= 1.02\times1014\text{J}.$$

Strategy for (b)

The efficiency can be calculated with $Eff = W / Q_h$ since Q_h is given and work W was found in the first part of this example.

Solution for (b)

Efficiency is given by: $Eff = W / Q_h$. The work W was just found to be 1.02×10^{14}J, and Q_h is given, so the efficiency is

$$Eff = (1.02\times10^{14}\text{J}) / (2.50\times10^{14}\text{J})$$

$$= 0.408, \text{ or } 40.8\%$$

Strategy for (c)

The daily consumption of coal is calculated using the information that each day there is 2.50×10^{14}J of heat transfer from coal. In the combustion process, we have $C + O_2 \Rightarrow CO_2$. So, every 12 kg of coal puts 12 kg + 16 kg + 16 kg = 44 kg of CO_2 into the atmosphere.

Solution for (c)

The daily coal consumption is

$$(2.50\times10^{14}\text{J}) / (2.50\times10^{6}\text{J/kg}) = 1.0\times10^{8}\text{kg}.$$

Assuming that the coal is pure and that all the coal goes toward producing carbon dioxide, the carbon dioxide produced per day is

$$1.0\times10^{8}\text{kg coal} \times (44 \text{ kg } CO_2) / (12 \text{ kg coal}) = 3.7\times10^{8} \text{ kg } CO_2$$

This is 370,000 metric tons of CO_2 produced every day.

Discussion

If all the work output is converted to electricity in a period of one day, the average power output is 1180 MW (this is left to you as an end-of-chapter problem). This value is about the size of a large-scale conventional power plant. The efficiency found is acceptably close to the value of 42% given for coal power stations. It means that fully 59.2% of the energy is heat transfer to the environment, which usually results in warming lakes, rivers, or the ocean near the power station, and is implicated in a warming planet generally. While the laws of thermodynamics limit the efficiency of such plants—including plants fired by nuclear fuel, oil, and natural gas—the heat transfer to the environment could be, and sometimes is, used for heating homes or for industrial processes. The generally low cost of energy has not made it economical to make better use of the waste heat transfer from most heat engines. Coal-fired power plants produce the greatest amount of CO_2 per unit energy output (compared to natural gas or oil), making coal the least efficient fossil fuel.

With the information given in the above example, we can find characteristics such as the efficiency of a heat engine without any knowledge of how the heat engine operates, but looking further into the mechanism of the engine will give us greater insight. Figure 15.13 illustrates the operation of the common four-stroke gasoline engine. The four steps shown complete this heat engine's cycle, bringing the gasoline-air mixture back to its original condition.

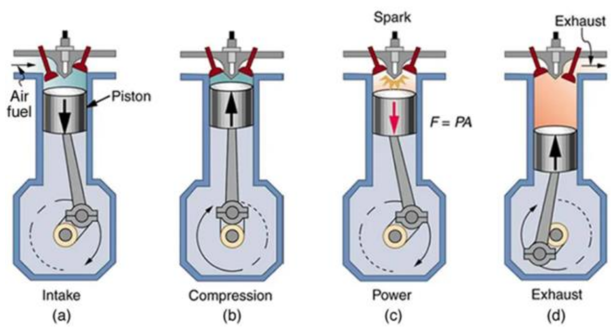


Figure **15.13:** In the four-stroke internal combustion gasoline engine, heat transfer into work takes place in the cyclical process shown here. The piston is connected to a rotating crankshaft, which both takes work out of and does work on the gas in the cylinder. (a) Air is mixed with fuel during the intake stroke. (b) During the compression stroke, the air-fuel mixture is rapidly compressed in a nearly adiabatic process, as the piston rises with the valves closed. Work is done on the gas. (c) The power stroke has two distinct parts. First, the air-fuel mixture is ignited, converting chemical potential energy into thermal energy almost instantaneously, which leads to a great increase in pressure. Then the piston descends, and the gas does work by exerting a force through a distance in a nearly adiabatic process. (d) The exhaust stroke expels the hot gas to prepare the engine for another cycle, starting again with the intake stroke.

The Otto cycle shown in Figure 15. 14 (a) is used in four-stroke internal combustion engines, although in fact the true Otto cycle paths do not correspond exactly to the strokes of the engine.

The adiabatic process AB corresponds to the nearly adiabatic compression stroke of the gasoline engine. In both cases, work is done on the system (the gas mixture in the cylinder), increasing its temperature and pressure. Along path BC of the Otto cycle, heat transfer Q_h into the gas occurs at constant volume, causing a further increase in pressure and temperature. This process corresponds to burning fuel in an internal combustion engine and takes place so rapidly that the volume is nearly constant. Path CD in the Otto cycle is an adiabatic expansion that does work on the outside world, just as the power stroke of an internal combustion engine does in its nearly adiabatic expansion. The work done by the system along path CD is greater than the work done on the system along path AB, because the pressure is greater, and so there is a net work output. Along path DA in the Otto cycle, heat transfer Q_c from the gas at constant volume reduces its temperature and pressure, returning it to its original state. In an internal combustion engine, this process corresponds to the exhaust of hot gases and the intake of an air-gasoline mixture at a considerably lower temperature. In both cases, heat transfer into the environment occurs along this final path.

The net work done by a cyclical process is the area inside the closed path on a PV diagram, such as that inside path ABCDA in Figure 15. 14. Note that in every imaginable cyclical process, it is absolutely necessary for heat transfer from the system to occur in order to get a net work output. In the Otto cycle, heat transfer occurs along path DA. If no heat transfer occurs, then the return path is the same, and the net work output is zero. The lower the temperature on the path AB, the less work has to be done to compress the gas. The area inside the closed path is then greater, and so the engine does more work and is thus more efficient. Similarly, the higher the temperature along path CD, the more work output there is. (See Figure 15.15) So efficiency is related to the temperatures of the hot and cold reservoirs. In the next section, we shall see what the absolute limit to the efficiency of a heat engine is, and how it is related to temperature.

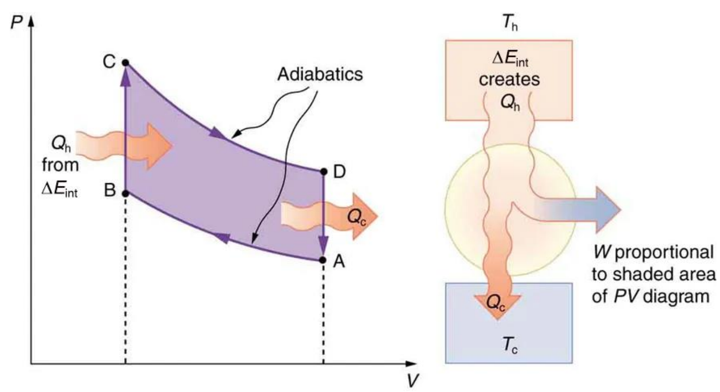


Figure **15.14:** PVPV diagram for a simplified Otto cycle, analogous to that employed in an internal combustion engine. Point A corresponds to the start of the compression stroke of an internal combustion engine. Paths AB and CD are adiabatic and correspond to the compression and power strokes of an internal combustion engine, respectively. Paths BC and DA are isochoric and accomplish similar results to the ignition and exhaust-intake portions, respectively, of the internal combustion engine's cycle. Work is done on the gas along path AB, but more work is done by the gas along path CD, so that there is a net work output.

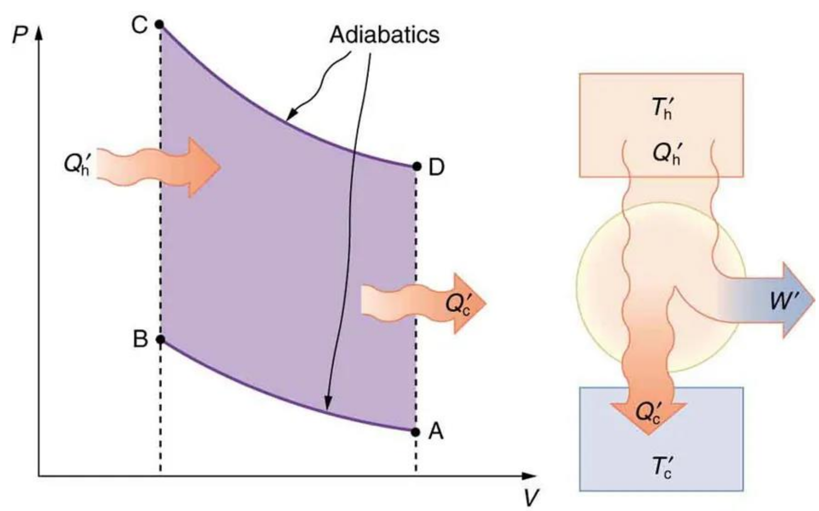


Figure **15.15:** This Otto cycle produces a greater work output than the one in Figure 15.14, because the starting temperature of path CD is higher and the starting temperature of path AB is lower. The area inside the loop is greater, corresponding to greater net work output.

We know from the second law of thermodynamics that a heat engine cannot be 100% efficient, since there must always be some heat transfer Q_c to the environment, which is often called waste heat. How efficient, then, can a heat engine be? This question was answered at a theoretical level in 1824 by a young French engineer, Sadi Carnot (1796–1832), in his study of the then-emerging heat engine technology crucial to the Industrial Revolution. He devised a theoretical cycle, now called the Carnot cycle, which is the most efficient cyclical process possible. The second law of thermodynamics can be restated in terms of the Carnot cycle, and so what Carnot actually discovered was this fundamental law. Any heat engine employing the Carnot cycle is called a Carnot engine.

What is crucial to the Carnot cycle—and, in fact, defines it—is that only reversible processes are used. Irreversible processes involve dissipative factors, such as friction and turbulence. This increases heat transfer Q_c to the environment and reduces the efficiency of the engine. Obviously, then, reversible processes are superior.

15. 8 Carnot Engine

Stated in terms of reversible processes, the second law of thermodynamics has a third form:

A Carnot engine operating between two given temperatures has the greatest possible efficiency of any heat engine operating between these two temperatures. Furthermore, all engines employing only reversible processes have this same maximum efficiency when operating between the same given temperatures.

Figure 15.16 shows the *PV* diagram for a Carnot cycle. The cycle comprises two isothermal and two adiabatic processes. Recall that both isothermal and adiabatic processes are, in principle, reversible.

Carnot also determined the efficiency of a perfect heat engine—that is, a Carnot engine. It is always true that the efficiency of a cyclical heat engine is given by:

$$Eff = (Q_h - Q_c) / (Q_h) = 1 - (Q_c/Q_h)$$

What Carnot found was that for a perfect heat engine, the ratio Q_c/Q_h equals the ratio of the absolute temperatures of the heat reservoirs. That is, $Q_c/Q_h = T_c/T_h$ for a Carnot engine, so that the maximum or Carnot efficiency Eff_C is given by

$$Eff_C = 1 - (T_c/T_h)$$

where T_h and T_c are in kelvins (or any other absolute temperature scale). No real heat engine can do as well as the Carnot efficiency. But the ideal Carnot engine, like the drinking bird above, while a fascinating novelty, has zero power. This makes it unrealistic for any applications.

Carnot’s interesting result implies that 100% efficiency would be possible only if $T_c = 0$ K — that is, only if the cold reservoir were at absolute zero, a practical and theoretical impossibility.

But the physical implication is this—the only way to have all heat transfer go into doing work is to remove *all* thermal energy, and this requires a cold reservoir at absolute zero.

It is also apparent that the greatest efficiencies are obtained when the ratio T_c/T_h is as small as possible. Just as discussed for the Otto cycle in the previous section, this means that efficiency is greatest for the highest possible temperature of the hot reservoir and lowest possible temperature of the cold reservoir. (This setup increases the area inside the closed loop on the *PV* diagram; also, it seems reasonable that the greater the temperature difference, the easier it is to divert the heat transfer to work.) The actual reservoir temperatures of a heat engine are usually related to the type of heat source and the temperature of the environment into which heat transfer occurs. Consider the following example.

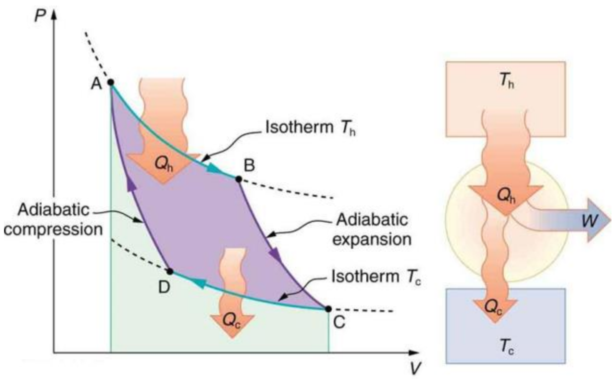


Figure 15.16: *PV* diagram for a Carnot cycle, employing only reversible isothermal and adiabatic processes. Heat transfer Q_h occurs into the working substance during the isothermal path AB, which takes place at constant temperature T_h. Heat transfer Q_c occurs out of the working substance during the isothermal path CD, which takes place at constant temperature T_c. The net work output W equals the area inside the path ABCDA. Also shown is a schematic of a Carnot engine operating between hot and cold reservoirs at temperatures T_h and T_c. Any heat engine using reversible processes and operating between these two temperatures will have the same maximum efficiency as the Carnot engine.

Example - Maximum Theoretical Efficiency for a Nuclear Reactor

A nuclear power reactor has pressurized water at 300ºC. (Higher temperatures are theoretically possible but practically not, due to limitations with materials used in the reactor.) Heat transfer from this water is a complex process (see Figure 15.17). Steam, produced in the steam generator, is used to drive the turbine-generators. Eventually the steam is condensed to water at 27ºC and then heated again to start the cycle over. Calculate the maximum theoretical efficiency for a heat engine operating between these two temperatures.

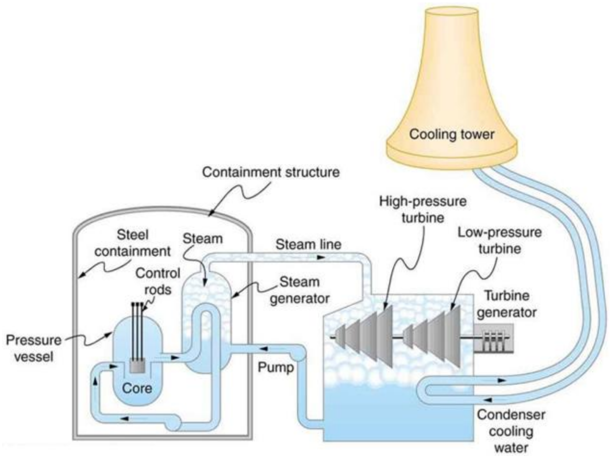


Figure 15.17: Schematic diagram of a pressurized water nuclear reactor and the steam turbines that convert work into electrical energy. Heat exchange is used to generate steam, in part to avoid contamination of the generators with radioactivity. Two turbines are used because this is less expensive than operating a single generator that produces the same amount of electrical energy. The steam is condensed to liquid before being returned to the heat exchanger, to keep exit steam pressure low and aid the flow of steam through the turbines (equivalent to using a lower-temperature cold reservoir). The considerable energy associated with condensation must be dissipated into the local environment; in this example, a cooling tower is used so there is no direct heat transfer to an aquatic environment. (Note that the water going to the cooling tower does not come into contact with the steam flowing over the turbines.)

Strategy

Since temperatures are given for the hot and cold reservoirs of this heat engine, $Eff_C = 1 - (T_c/T_h)$ can be used to calculate the Carnot (maximum theoretical) efficiency. Those temperatures must first be converted to kelvins.

Solution

The hot and cold reservoir temperatures are given as 300ºC and 27.0ºC, respectively. In kelvins, then, T_h = 573 K and T_c = 300 K, so that the maximum efficiency is

$$Eff_C = 1 - (T_c/T_h)$$

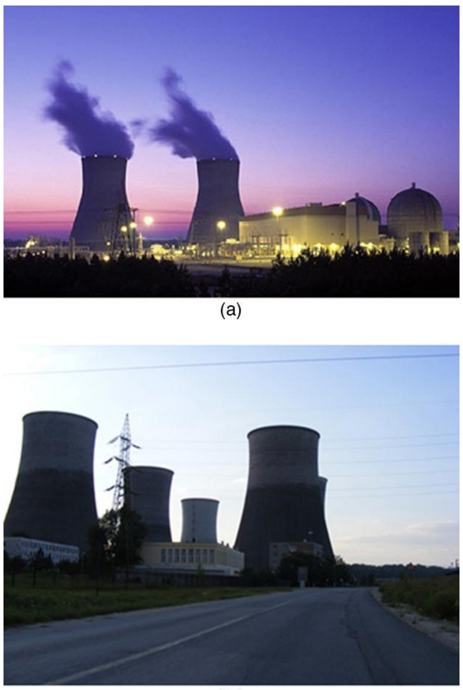
Thus,

$$Eff_C = 1 - (300\text{ K}/573\text{ K})$$

$$= 0.476\text{, or }47.6\%.$$

Discussion

A typical nuclear power station's actual efficiency is about 35%, a little better than 0.7 times the maximum possible value, a tribute to superior engineering. Electrical power stations fired by coal, oil, and natural gas have greater actual efficiencies (about 42%), because their boilers can reach higher temperatures and pressures. The cold reservoir temperature in any of these power stations is limited by the local environment. Figure 15.18 shows (a) the exterior of a nuclear power station and (b) the exterior of a coal-fired power station. Both have cooling towers into which water from the condenser enters the tower near the top and is sprayed downward, cooled by evaporation.



(a)

(b)

Figure 15.18: (a) A nuclear power station (credit: BlatantWorld.com) and (b) a coal-fired power station. Both have cooling towers in which water evaporates into the environment, representing Q_c. The nuclear reactor, which supplies Q_h, is housed inside the dome-shaped containment buildings. (credit: Robert & Mihaela Vicol, publicphoto.org)

Since all real processes are irreversible, the actual efficiency of a heat engine can never be as great as that of a Carnot engine, as illustrated in Figure 15.19 (a). Even with the best heat engine possible, there are always dissipative processes in peripheral equipment, such as electrical transformers or car transmissions. These further reduce the overall efficiency by converting some of the engine's work output back into heat transfer, as shown in Figure 15.19 (b).

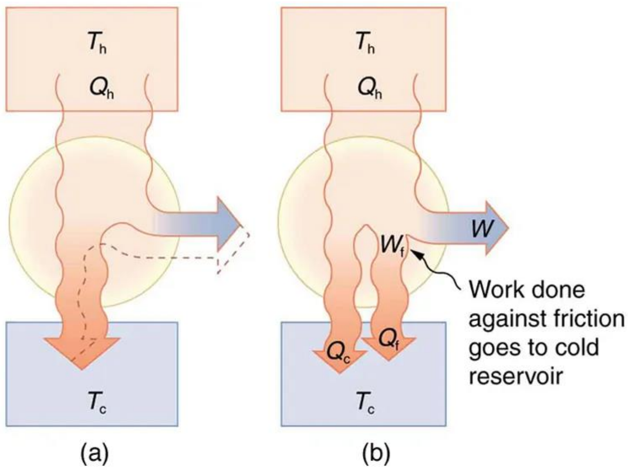


Figure **15.19:** Real heat engines are less efficient than Carnot engines. (a) Real engines use irreversible processes, reducing the heat transfer to work. Solid lines represent the actual process; the dashed lines are what a Carnot engine would do between the same two reservoirs. (b) Friction and other dissipative processes in the output mechanisms of a heat engine convert some of its work output into heat transfer to the environment.

Heat pumps, air conditioners, and refrigerators utilize heat transfer from cold to hot. They are heat engines run backward. We say backward, rather than reverse, because except for Carnot engines, all heat engines, though they can be run backward, cannot truly be reversed. Heat transfer occurs from a cold reservoir Q_c and into a hot one. This requires work input W, which is also converted to heat transfer. Thus the heat transfer to the hot reservoir is $Q_h = Q_c + W$. (Note that Q_h, Q_c, and W are positive, with their directions indicated on schematics rather than by sign.) A heat pump's mission is for heat transfer Q_h to occur into a warm environment, such as a home in the winter. The mission of air conditioners and refrigerators is for heat transfer Q_c to occur from a cool environment, such as chilling a room or keeping food at lower temperatures than the environment. (Actually, a heat pump can be used both to heat and cool a space. It is essentially an air conditioner and a heating unit all in one. In this section we will concentrate on its heating mode.)

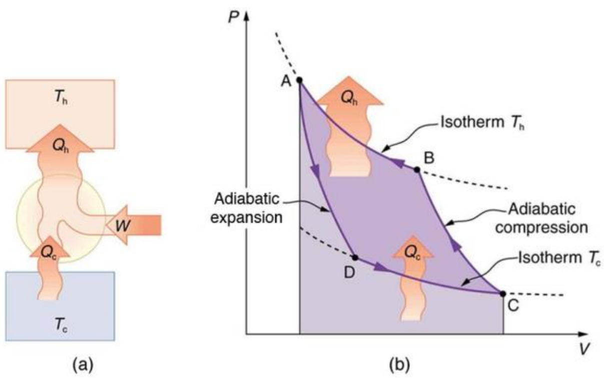


Figure 15.20: Heat pumps, air conditioners, and refrigerators are heat engines operated backward. The one shown here is based on a Carnot (reversible) engine. (a) Schematic diagram showing heat transfer from a cold reservoir to a warm reservoir with a heat pump. The directions of ***W***, *Q*h, and *Q*c are opposite what they would be in a heat engine. (b) *PV* diagram for a Carnot cycle similar to that in Figure 15.21 but reversed, following path ADCBA. The area inside the loop is negative, meaning there is a net work input. There is heat transfer *Q*c into the system from a cold reservoir along path DC, and heat transfer *Q*h out of the system into a hot reservoir along path BA.

15.9 Heat Pumps

The great advantage of using a heat pump to keep your home warm, rather than just burning fuel, is that a heat pump supplies $Qh = Qc + W$. Heat transfer is from the outside air, even at a temperature below freezing, to the indoor space. You only pay for *W*, and you get an additional heat transfer of *Q*c from the outside at no cost; in many cases, at least twice as much energy is transferred to the heated space as is used to run the heat pump. When you burn fuel to keep warm, you pay for all of it. The disadvantage is that the work input (required by the second law of thermodynamics) is sometimes more expensive than simply burning fuel, especially if the work is done by electrical energy.

The basic components of a heat pump in its heating mode are shown in Figure 15.21. A working fluid such as a non-CFC refrigerant is used. In the outdoor coils (the evaporator), heat transfer *Q*c occurs to the working fluid from the cold outdoor air, turning it into a gas.

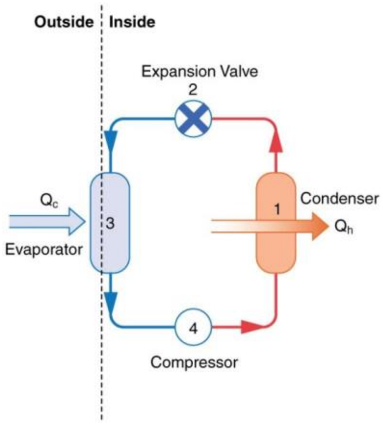


Figure 15.21: A simple heat pump has four basic components: (1) condenser, (2) expansion valve, (3) evaporator, and (4) compressor. In the heating mode, heat transfer Q_c occurs to the working fluid in the evaporator (3) from the colder outdoor air, turning it into a gas. The electrically driven compressor (4) increases the temperature and pressure of the gas and forces it into the condenser coils (1) inside the heated space. Because the temperature of the gas is higher than the temperature in the room, heat transfer from the gas to the room occurs as the gas condenses to a liquid. The working fluid is then cooled as it flows back through an expansion valve (2) to the outdoor evaporator coils.

The electrically driven compressor (work input W) raises the temperature and pressure of the gas and forces it into the condenser coils that are inside the heated space. Because the temperature of the gas is higher than the temperature inside the room, heat transfer to the room occurs and the gas condenses to a liquid. The liquid then flows back through a pressure-reducing valve to the outdoor evaporator coils, being cooled through expansion. (In a cooling cycle, the evaporator and condenser coils exchange roles and the flow direction of the fluid is reversed.)

The quality of a heat pump is judged by how much heat transfer Q_h occurs into the warm space compared with how much work input W is required. In the spirit of taking the ratio of what you get to what you spend, we define a heat pump's coefficient of performance (COP_{hp}) to be

$$COP_{hp} = Q_h / W.$$

Since the efficiency of a heat engine is $Eff = W/Q_h$, we see that $COP_{hp} = 1/Eff$, an important and interesting fact. First, since the efficiency of any heat engine is less than 1, it means that COP_{hp} is always greater than 1 — that is, a heat pump always has more heat transfer Q_h than work put into it. Second, it means that heat pumps work best when temperature differences are small. The efficiency of a perfect, or Carnot, engine is $Eff_C = 1-(T_c/T_h)$; thus, the smaller the temperature difference, the smaller the efficiency and the greater the COP_{hp} (because $COP_{hp}=1/Eff$). In other words, heat pumps do not work as well in very cold climates as they do in more moderate climates.

Friction and other irreversible processes reduce heat engine efficiency, but they do *not* benefit the operation of a heat pump—instead, they reduce the work input by converting part of it to heat transfer back into the cold reservoir before it gets into the heat pump.

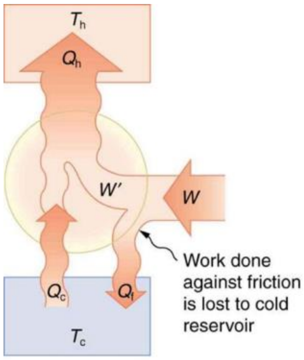


Figure 15.22: When a real heat engine is run backward, some of the intended work input (*W*) goes into heat transfer before it gets into the heat engine, thereby reducing its coefficient of performance *COP*hp. In this figure, *W′* represents the portion of *W* that goes into the heat pump, while the remainder of *W* is lost in the form of frictional heat (*Qf*) to the cold reservoir. If all of ***W*** had gone into the heat pump, then *Q*h would have been greater. The best heat pump uses adiabatic and isothermal processes, since, in theory, there would be no dissipative processes to reduce the heat transfer to the hot reservoir.

Example - The Best *COP* $_{\mathbf{hp}}$ of a Heat Pump for Home Use

A heat pump used to warm a home must employ a cycle that produces a working fluid at temperatures greater than typical indoor temperature so that heat transfer to the inside can take place. Similarly, it must produce a working fluid at temperatures that are colder than the outdoor temperature so that heat transfer occurs from outside. It is hot and cold reservoir temperatures; therefore cannot be too close, placing a limit on its *COP*hp. (See Figure 15.23.) What is the best coefficient of performance possible for such a heat pump, if it has a hot reservoir temperature of 45.0ºC and a cold reservoir temperature of −15.0ºC?

Strategy

A Carnot engine reversed will give the best possible performance as a heat pump. As noted above, *COP*hp=1/*Eff*, so that we need to first calculate the Carnot efficiency to solve this problem.

Solution

Carnot efficiency in terms of absolute temperature is given by**:**

$$Eff_C = 1 – (Tc/Th)$$

The temperatures in kelvins are Th=318 K and Tc=258 K, so that

$$Eff_C = 1 – (258\ \text{K}/318\ \text{K}) = 0.1887.$$

Thus, from the discussion above,

$$COP_{hp} = 1 / Eff = 1/0.1887 = 5.30,$$

or

$$COP_{hp} = Qh/W = 5.30$$

so that

$$Qh = 5.30\ \text{W}$$

Discussion

This result means that the heat transfer by the heat pump is 5.30 times as much as the work put into it. It would cost 5.30 times as much for the same heat transfer by an electric room heater as it does for that produced by this heat pump. This is not a violation of conservation of energy. Cold ambient air provides 4.3 J per 1 J of work from the electrical outlet.

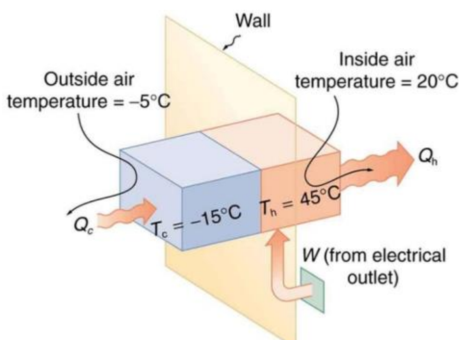


Figure 15.23: Heat transfer from the outside to the inside, along with work done to run the pump, takes place in the heat pump of the example above**.** Note that the cold temperature produced by the heat pump is lower than the outside temperature, so that heat transfer into the working fluid occurs. The pump’s compressor produces a temperature greater than the indoor temperature in order for heat transfer into the house to occur.

Real heat pumps do not perform quite as well as the ideal one in the previous example; their values of *COP*hp range from about 2 to 4. This range means that the heat transfer *Q*h from the heat pumps is 2 to 4 times as great as the work *W* put into them. Their economical feasibility is still limited, however, since W is usually supplied by electrical energy that costs more per joule than heat transfer by burning fuels like natural gas. Furthermore, the initial cost of a heat pump is greater than that of many furnaces, so that a heat pump must last longer for its cost to be recovered. Heat pumps are most likely to be economically superior where winter temperatures are mild, electricity is relatively cheap, and other fuels are relatively expensive. Also, since they can cool as well as heat a space, they have advantages where cooling in summer months is also desired. Thus some of the best locations for heat pumps are in warm summer climates with cool winters. Figure 15.24 shows a heat pump, called a "*reverse cycle"* or "*split-system cooler"* in some countries.

Figure 15.24 In hot weather, heat transfer occurs from air inside the room to air outside, cooling the room. In cool weather, heat transfer occurs from air outside to air inside, warming the room. This switching is achieved by reversing the direction of flow of the working fluid.

15.10 Air Conditioners and Refrigerators

Air conditioners and refrigerators are designed to cool something down in a warm environment. As with heat pumps, work input is required for heat transfer from cold to hot, and this is expensive. The quality of air conditioners and refrigerators is judged by how much heat transfer Q_c occurs from a cold environment compared with how much work input W is required. What is considered the benefit in a heat pump is considered waste heat in a refrigerator. We thus define the coefficient of performance (COP_{ref}) of an air conditioner or refrigerator to be

$$COP_{ref} = Q_c / W$$

Noting again that $Q_h = Q_c + W$, we can see that an air conditioner will have a lower coefficient of performance than a heat pump, because $COP_{hp} = Q_h/W$ and Q_h is greater than Q_c. In this module's Problems and Exercises, you will show that

$$COP_{ref} = COP_{hp} - 1$$

for a heat engine used as either an air conditioner or a heat pump operating between the same two temperatures. Real air conditioners and refrigerators typically do remarkably well, having values of COP_{ref} ranging from 2 to 6. These numbers are better than the COP_{hp} values for the heat pumps mentioned above, because the temperature differences are smaller, but they are less than those for Carnot engines operating between the same two temperatures.

A type of *COP* rating system called the “energy efficiency rating” (*EER*) has been developed. This rating is an example where non-SI units are still used and relevant to consumers. To make it easier for the consumer, Australia, Canada, New Zealand, and the U.S. use an Energy Star Rating out of 5 stars—the more stars, the more energy efficient the appliance. *EER*s are expressed in mixed units of British thermal units (Btu) per hour of heating or cooling divided by the power input in watts. Room air conditioners are readily available with *EER*s ranging from 6 to 12. Although not the same as the *COP*s just described, these *EER*s are good for comparison purposes—the greater the *EER*, the cheaper an air conditioner is to operate (but the higher its purchase price is likely to be).

The *EER* of an air conditioner or refrigerator can be expressed as

$$EER = (Q_c/t_1) / (W/t_2)$$

where Q_c is the amount of heat transfer from a cold environment in British thermal units, t_1 is time in hours, W is the work input in joules, and t_2 is time in seconds.

Recall that the simple definition of energy is the ability to do work. Entropy is a measure of how much energy is not available to do work. Although all forms of energy are interconvertible, and all can be used to do work, it is not always possible, even in principle, to convert the entire available energy into work. That unavailable energy is of interest in thermodynamics, because the field of thermodynamics arose from efforts to convert heat to work.

We can see how entropy is defined by recalling our discussion of the Carnot engine. We noted that for a Carnot cycle, and hence for any reversible processes, $Q_c/Q_h = T_c/T_h$. Rearranging terms yields

$$Q_c / T_c = Q_h / T_h$$

for any reversible process. Q_c and Q_h are absolute values of the heat transfer at temperatures T_c and T_h, respectively. This ratio of Q/T is defined to be the change in entropy ΔS for a reversible process,

$$\Delta S = (Q / T)_{\text{rev}}$$

where Q is the heat transfer, which is positive for heat transfer into and negative for heat transfer out of, and T is the absolute temperature at which the reversible process takes place. The SI unit for entropy is joules per kelvin (J/K). If temperature changes during the process, then it is usually a good approximation (for small changes in temperature) to take T to be the average temperature, avoiding the need to use integral calculus to find ΔS.

The definition of ΔS is strictly valid only for reversible processes, such as used in a Carnot engine. However, we can find ΔS precisely even for real, irreversible processes. The reason is that the entropy *s* of a system, like internal energy *U*, depends only on the state of the system and not how it reached that condition. Entropy is a property of state. Thus, the change in entropy ΔS of a system between state 1 and state 2 is the same no matter how the change occurs. We just need to find or imagine a reversible process that takes us from state 1 to state 2 and calculate ΔS for that process. That will be the change in entropy for any process going from state 1 to state 2. (See Figure 15.25)

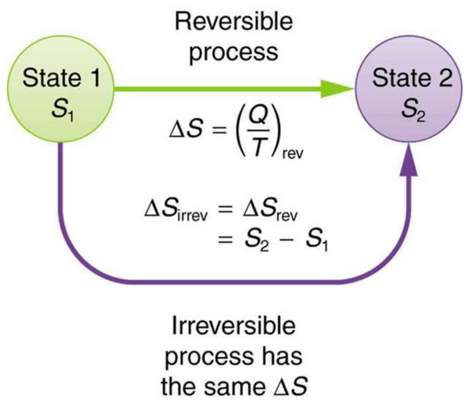


Figure 15.25: When a system goes from state 1 to state 2, its entropy changes by the same amount ΔS, whether a hypothetical reversible path is followed or a real irreversible path is taken.

Now let us take a look at the change in entropy of a Carnot engine and its heat reservoirs for one full cycle. The hot reservoir has a loss of entropy $\Delta S_h = -(Qh/Th)$, because heat transfer occurs out of it (remember that when heat transfers out, then *Q* has a negative sign). The cold reservoir has a gain of entropy $\Delta S_c = (Qc/Tc)$, because heat transfer occurs into it. (We assume the reservoirs are sufficiently large that their temperatures are constant.) So, the total change in entropy is

$$\Delta S_{tot} = \Delta S_h + \Delta S_c$$

Thus, since we know that $Qh/Th = Qc/Tc$ for a Carnot engine,

$$\Delta S_{tot} = -(Qh / Th) + (Qc/Tc) = 0$$

This result, which has general validity, means that *the total change in entropy for a system in any reversible process is zero.*

The entropy of various parts of the system may change, but the total change is zero. Furthermore, the system does not affect the entropy of its surroundings since heat transfer between them does not occur. Thus, the reversible process changes neither the total entropy of the system nor the entropy of its surroundings. Sometimes this is stated as follows: *Reversible processes do not affect the total entropy of the universe.* Real processes are not reversible, though, and they do change total entropy. We can, however, use hypothetical reversible processes to determine the value of entropy in real, irreversible processes. The following example illustrates this point.

Example - Entropy Increases in an Irreversible (Real) Process

Spontaneous heat transfer from hot to cold is an irreversible process. Calculate the total change in entropy if 4000 J of heat transfer occurs from a hot reservoir at T_h= 600 K(327º C) to a cold reservoir at T_c = 250 K(−23º C), assuming there is no temperature change in either reservoir. (See Figure 15.26)

Strategy

How can we calculate the change in entropy for an irreversible process when $\Delta S_{tot} = \Delta S_h + \Delta S_c$ is valid only for reversible processes? Remember that the total change in entropy of the hot and cold reservoirs will be the same whether a reversible or irreversible process is involved in heat transfer from hot to cold. So, we can calculate the change in entropy of the hot reservoir for a hypothetical reversible process in which 4000 J of heat transfer occurs from it; then we do the same for a hypothetical reversible process in which 4000 J of heat transfer occurs to the cold reservoir. This produces the same changes in the hot and cold reservoirs that would occur if the heat transfer were allowed to occur irreversibly between them, and so it also produces the same changes in entropy.

Solution

We now calculate the two changes in entropy using $\Delta S_{tot} = \Delta S_h + \Delta S_c$. First, for the heat transfer from the hot reservoir,

$$\Delta S_h = -\left(Qh/Th\right) = -(4000\ \text{J}/600\ \text{K}) = -6.67\ \text{J/K}$$

And for the cold reservoir,

$$\Delta Sc = Qc\ /Tc = 4000\ \text{J}/\ 250\ \text{K} = 16.0\ \text{J/K}$$

Thus, the total is

$$\Delta S_{tot} = \ \Delta S_h + \Delta S_c\ (-6.67 + 16.0)\text{J/K}$$

$$= 9.33\ \text{J/K}$$

Discussion

There is an *increase* in entropy for the system of two heat reservoirs undergoing this irreversible heat transfer. We will see that this means there is a loss of ability to do work with this transferred energy. Entropy has increased, and energy has become unavailable to do work.

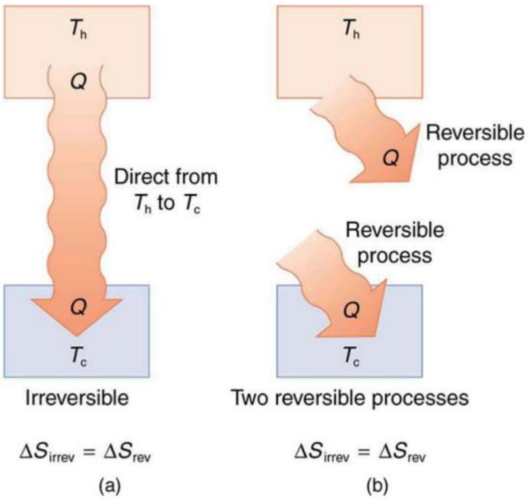


Figure 15.26: (a) Heat transfer from a hot object to a cold one is an irreversible process that produces an overall increase in entropy. (b) The same final state and, thus, the same change in entropy is achieved for the objects if reversible heat transfer processes occur between the two objects whose temperatures are the same as the temperatures of the corresponding objects in the irreversible process.

It is reasonable that entropy increases for heat transfer from hot to cold. Since the change in entropy is Q/T, there is a larger change at lower temperatures. The decrease in entropy of the hot object is therefore less than the increase in entropy of the cold object, producing an overall increase, just as in the previous example. This result is very general:

There is an increase in entropy for any system undergoing an irreversible process.

With respect to entropy, there are only two possibilities: entropy is constant for a reversible process, and it increases for an irreversible process. There is a fourth version of the second law of thermodynamics stated in terms of entropy:

The total entropy of a system either increases or remains constant in any process; it never decreases.

For example, heat transfer cannot occur spontaneously from cold to hot, because entropy would decrease.

Entropy is very different from energy. Entropy is *not* conserved but increases in all real processes. Reversible processes (such as in Carnot engines) are the processes in which the most heat transfer to work takes place and are also the ones that keep entropy constant. Thus, we are led to make a connection between entropy and the availability of energy to do work.

Entropy and the Unavailability of Energy to Do Work

What does a change in entropy mean, and why should we be interested in it? One reason is that entropy is directly related to the fact that not all heat transfer can be converted into work. The next example gives some indication of how an increase in entropy results in less heat transfer into work.

Example - Less Work is Produced by a Given Heat Transfer When Entropy Change is Greater

(a)Calculate the work output of a Carnot engine operating between temperatures of 600 K and 100 K for 4000 J of heat transfer to the engine. (b) Now suppose that the 4000 J of heat transfer occurs first from the 600 K reservoir to a 250 K reservoir (without doing any work, and this produces the increase in entropy calculated above) before transferring into a Carnot engine operating between 250 K and 100 K. What work output is produced? (See Figure 15.27)

Strategy

In both parts, we must first calculate the Carnot efficiency and then the work output.

Solution (a)

The Carnot efficiency is given by

$$Eff_C = 1 - (T_c/T_h)$$

Substituting the given temperatures yields

$$Eff_C = 1 - (100\text{ K} / 600\text{ K}) = 0.833$$

Now the work output can be calculated using the definition of efficiency for any heat engine as given by

$$Eff = W / Q_h$$

Solving for *W* and substituting known terms gives

$$W = Eff_C / Q_h$$

$$= (0.833)\,(4000\text{J}) = 3333\text{ J}$$

Solution (b)

Similarly,

$$Eff'_C = 1 - (T_C/T'_C) = 1 - (100\text{ K}/250\text{ K}) = 0.600$$

so that

$$W = Eff'_C Q_h$$

$$= (0.600)(4000 \text{ J}) = 2400 \text{ J}$$

Discussion

There is 933 J less work from the same heat transfer in the second process. This result is important. The same heat transfer into two perfect engines produces different work outputs, because the entropy change differs in the two cases. In the second case, entropy is greater and less work is produced. Entropy is associated with the *un*availability of energy to do work.

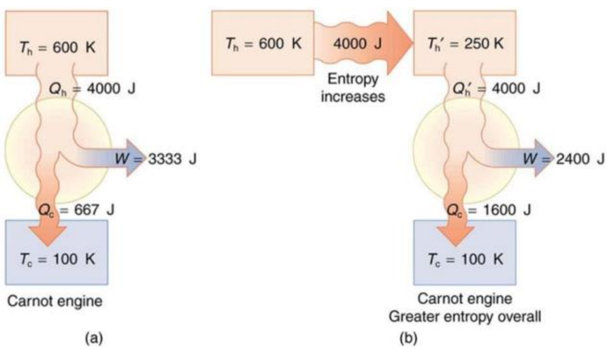


Figure 15.27: (a) A Carnot engine working at between 600 K and 100 K has 4000 J of heat transfer and performs 3333 J of work. (b) The 4000 J of heat transfer occurs first irreversibly to a 250 K reservoir and then goes into a Carnot engine. The increase in entropy caused by the heat transfer to a colder reservoir results in a smaller work output of 2400 J. There is a permanent loss of 933 J of energy for the purpose of doing work.

When entropy increases, a certain amount of energy becomes *permanently* unavailable to do work. The energy is not lost, but its character is changed, so that some of it can never be converted to doing work—that is, to an organized force acting through a distance. For instance, in the previous example, 933 J less work was done after an increase in entropy of 9.33 J/K occurred in the 4000 J heat transfer from the 600 K reservoir to the 250 K reservoir. It can be shown that the amount of energy that becomes unavailable for work is

$$W_{\text{unavail}} = \Delta S \cdot T_0$$

where T_0 is the lowest temperature utilized. In the previous example,

$$W_{\text{unavai l}} = (9.33 \text{ J/K})(100 \text{ K}) = 933 \text{ J}$$

as found.

Heat Death of the Universe: An Overdose of Entropy

In the early, energetic universe, all matter and energy were easily interchangeable and identical in nature. Gravity played a vital role in the young universe. Although it may have *seemed* disorderly, and therefore, superficially entropic, in fact, there was enormous potential energy available to do work—all the future energy in the universe.

As the universe matured, temperature differences arose, which created more opportunity for work. Stars are hotter than planets, for example, which are warmer than icy asteroids, which are warmer still than the vacuum of the space between them.

Most of these are cooling down from their usually violent births, at which time they were provided with energy of their own—nuclear energy in the case of stars, volcanic energy on Earth and other planets, and so on. Without additional energy input, however, their days are numbered.

As entropy increases, less and less energy in the universe is available to do work. On Earth, we still have great stores of energy such as fossil and nuclear fuels; large-scale temperature differences, which can provide wind energy; geothermal energies due to differences in temperature in Earth's layers; and tidal energies owing to our abundance of liquid water. As these are used, a certain fraction of the energy they contain can never be converted into doing work. Eventually, all fuels will be exhausted, all temperatures will equalize, and it will be impossible for heat engines to function, or for work to be done.

Entropy increases in a closed system, such as the universe. But in parts of the universe, for instance, in the Solar system, it is not a locally closed system. Energy flows from the Sun to the planets, replenishing Earth's stores of energy. The Sun will continue to supply us with energy for about another five billion years. We will enjoy direct solar energy, as well as side effects of solar energy, such as wind power and biomass energy from photosynthetic plants. The energy from the Sun will keep our water at the liquid state, and the Moon's gravitational pull will continue to provide tidal energy. But Earth's geothermal energy will slowly run down and won't be replenished.

But in terms of the universe, and the very long-term, very large-scale picture, the entropy of the universe is increasing, and so the availability of energy to do work is constantly decreasing. Eventually, when all stars have died, all forms of potential energy have been utilized, and all temperatures have equalized (depending on the mass of the universe, either at a very high temperature following a universal contraction, or a very low one, just before all activity ceases) there will be no possibility of doing work.

Either way, the universe is destined for thermodynamic equilibrium—maximum entropy. This is often called the *heat death of the universe*, and will mean the end of all activity. However, whether the universe contracts and heats up, or continues to expand and cools down, the end is not near. Calculations of black holes suggest that entropy can easily continue for at least 10^{100} years.

Order to Disorder

Entropy is related not only to the unavailability of energy to do work—it is also a measure of disorder. This notion was initially postulated by Ludwig Boltzmann in the 1800s. For example, melting a block of ice means taking a highly structured and orderly system of water molecules and converting it into a disorderly liquid in which molecules have no fixed positions (See Figure 15. 28). There is a large increase in entropy in the process, as seen in the following example.

Example - Entropy Associated with Disorder

Find the increase in entropy of 1.00 kg of ice originally at 0º C that is melted to form water at 0º C.

Strategy

As before, the change in entropy can be calculated from the definition of ΔS once we find the energy Q needed to melt the ice.

Solution

The change in entropy is defined as:

$$\Delta S = Q/T$$

Here Q is the heat transfer necessary to melt 1.00 kg of ice and is given by

$$Q = mL_f$$

where m is the mass and L_f is the latent heat of fusion. L_f = 334kJ/kg for water, so that

$$Q = (1.00 \text{ kg})(334 \text{ kJ/kg}) = 3.34\times10^5 \text{J}.$$

Now the change in entropy is positive, since heat transfer occurs into the ice to cause the phase change; thus,

$$\Delta S = Q / T = (3.34 \times 10^5 \text{J}) / T$$

T is the melting temperature of ice. That is, T = 0ºC = 273 K. So, the change in entropy is

$$\Delta S = (3.34\times10^5 \text{J}) / (273 \text{ K})$$

$$= 1.22\times10^3 \text{J/K}$$

Discussion

This is a significant increase in entropy accompanying an increase in disorder.

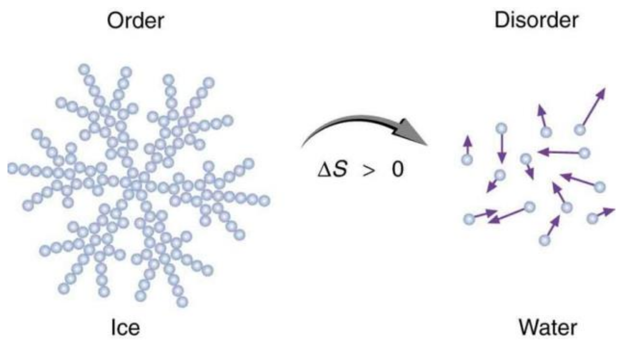


Figure 15.28: When ice melts, it becomes more disordered and less structured. The systematic arrangement of molecules in a crystal structure is replaced by a more random and less orderly movement of molecules without fixed locations or orientations. Its entropy increases because heat transfer occurs into it. Entropy is a measure of disorder.

In another easily imagined example, suppose we mix equal masses of water originally at two different temperatures, say 20.0º C and 40.0º C. The result is water at an intermediate temperature of 30.0º C. Three outcomes have resulted: entropy has increased, some energy has become unavailable to do work, and the system has become less orderly. Let us think about each of these results.

First, entropy has increased for the same reason that it did in the example above. Mixing the two bodies of water has the same effect as heat transfer from the hot one and the same heat transfer into the cold one. The mixing decreases the entropy of the hot water but increases the entropy of the cold water by a greater amount, producing an overall increase in entropy.

Second, once the two masses of water are mixed, there is only one temperature—you cannot run a heat engine with them. The energy that could have been used to run a heat engine is now unavailable to do work.

Third, the mixture is less orderly, or to use another term, less structured. Rather than having two masses at different temperatures and with different distributions of molecular speeds, we now have a single mass with a uniform temperature.

These three results—entropy, unavailability of energy, and disorder—are not only related but are in fact essentially equivalent.

Life, Evolution, and the Second Law of Thermodynamics

Some people misunderstand the second law of thermodynamics, stated in terms of entropy, to say that the process of the evolution of life violates this law. Over time, complex organisms evolved from much simpler ancestors, representing a large decrease in entropy of the Earth's biosphere. It is a fact that living organisms have evolved to be highly structured, and much lower in entropy than the substances from which they grow. But it is *always* possible for the entropy of one part of the universe to decrease, provided the total change in entropy of the universe increases. In equation form, we can write this as

$$\Delta S_{tot} = \Delta S_{syst} + \Delta S_{envir} > 0$$

Thus, ΔS_{syst} can be negative as long as ΔS_{envir} is positive and greater in magnitude.

How is it possible for a system to decrease its entropy? Energy transfer is necessary. If I pick up marbles that are scattered about the room and put them into a cup, my work has decreased the entropy of that system. If I gather iron ore from the ground and convert it into steel and build a bridge, my work has decreased the entropy of that system. Energy coming from the Sun can decrease the entropy of local systems on Earth—that is, ΔS_{syst} is negative. But the overall entropy of the rest of the universe increases by a greater amount—that is, ΔS_{envir} is positive and greater in magnitude. Thus, $\Delta S_{tot} = \Delta S_{syst} + \Delta S_{envir} > 0$, and the second law of thermodynamics is *not* violated.

Every time a plant stores some solar energy in the form of chemical potential energy, or an updraft of warm air lifts a soaring bird, the Earth can be viewed as a heat engine operating between a hot reservoir supplied by the Sun and a cold reservoir supplied by dark outer space—a heat engine of high complexity, causing local decreases in entropy as it uses part of the heat transfer from the Sun into deep space. There is a large total increase in entropy resulting from this massive heat transfer. A small part of this heat transfer is stored in structured systems on Earth, producing much smaller local decreases in entropy. (See Figure 15.29)

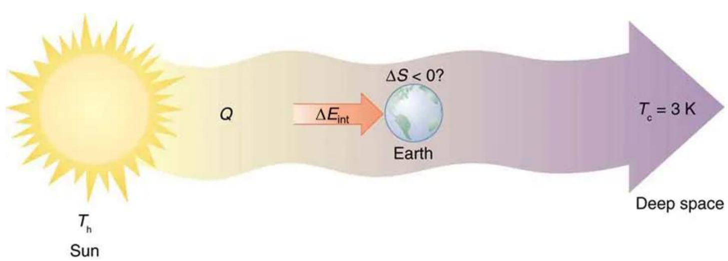


Figure 15.29: Earth's entropy may decrease in the process of intercepting a small part of the heat transfer from the Sun into deep space. Entropy for the entire process increases greatly while Earth becomes more structured with living systems and stored energy in various forms.

The various ways of formulating the second law of thermodynamics tell what happens rather than why it happens. Why should heat transfer occur only from hot to cold? Why should energy become ever less available to do work? Why should the universe become increasingly disorderly? The answer is that it is a matter of overwhelming probability. Disorder is simply vastly more likely than order.

When you watch an emerging rainstorm begin to wet the ground, you will notice that the drops fall in a disorganized manner both in time and in space. Some fall close together, some far apart, but they never fall in straight, orderly rows. It is not impossible for rain to fall in an orderly pattern, just highly unlikely, because there are many more disorderly ways than orderly ones. To illustrate this fact, we will examine some random processes, starting with coin tosses.

Disorder in a Gas

The fantastic growth in the odds favoring disorder that we see in going from 5 to 100 coins continues as the number of entities in the system increases. Let us now imagine applying this approach to perhaps a small sample of gas. Because counting microstates and macrostates involves statistics, this is called statistical analysis. The macrostates of a gas correspond to its macroscopic properties, such as volume, temperature, and pressure; and its microstates correspond to the detailed description of the positions and velocities of its atoms. Even a small amount of gas has a huge number of atoms: 1.0 cm^3 of an ideal gas at 1.0 atm and 0º C has 2.7×10^{19} atoms. So each macrostate has an immense number of microstates. In plain language, this means that there are an immense number of ways in which the atoms in a gas can be arranged, while still having the same pressure, temperature, and so on.

The most likely conditions (or macrostates) for a gas are those we see all the time—a random distribution of atoms in space with a Maxwell-Boltzmann distribution of speeds in random directions, as predicted by kinetic theory. This is the most disorderly and least structured condition we can imagine. In contrast, one type of very orderly and structured macrostate has all of the atoms in one corner of a container with identical velocities. There are very few ways to accomplish this (very few microstates corresponding to it), and so it is exceedingly unlikely ever to occur. (See Figure 15.30 (b).) Indeed, it is so unlikely that we have a law saying that it is impossible, which has never been observed to be violated—the second law of thermodynamics.

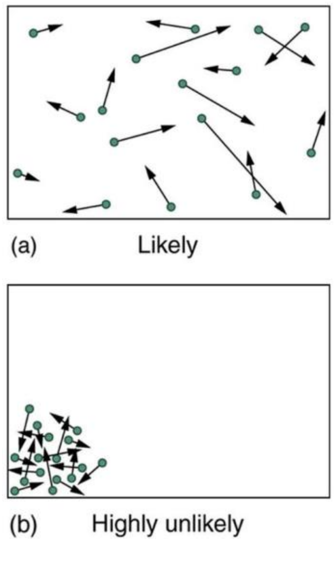


Figure 15.30: (a) The ordinary state of gas in a container is a disorderly, random distribution of atoms or molecules with a Maxwell-Boltzmann distribution of speeds. It is so unlikely that these atoms or molecules would ever end up in one corner of the container that it might as well be impossible. (b) With energy transfer, the gas can be forced into one corner and its entropy greatly reduced. But left alone, it will spontaneously increase its entropy and return to the normal conditions, because they are immensely more likely.

The disordered condition is one of high entropy, and the ordered one has low entropy. With a transfer of energy from another system, we could force all of the atoms into one corner and have a local decrease in entropy, but at the cost of an overall increase in entropy of the universe. If the atoms start out in one corner, they will quickly disperse and become uniformly distributed and will never return to the orderly original state Figure 15.30 (b)). Entropy will increase. With such a large sample of atoms, it is possible—but unimaginably unlikely—for entropy to decrease. Disorder is vastly more likely than order.

The arguments that disorder and high entropy are the most probable states are quite convincing. The great Austrian physicist Ludwig Boltzmann (1844–1906)—who, along with Maxwell, made so many contributions to kinetic theory—proved that the entropy of a system in a given state (a macrostate) can be written as

$$S = k \ln W$$

where $k = 1.38 \times 10^{-23 J}/\mathrm{K}$ is Boltzmann's constant, and $\ln W$ is the natural logarithm of the number of microstates W corresponding to the given macrostate. W is proportional to the probability that the macrostate will occur. Thus entropy is directly related to the probability of a state—the more likely the state, the greater its entropy. Boltzmann proved that this expression for S is equivalent to the definition $\Delta S = Q/T$, which we have used extensively.

Thus, the second law of thermodynamics is explained on a very basic level: entropy either remains the same or increases in every process. This phenomenon is due to the extraordinarily small probability of a decrease, based on the extraordinarily larger number of microstates in systems with greater entropy. Entropy *can* decrease, but for any macroscopic system, this outcome is so unlikely that it will never be observed.

For examples and answers, please refer to OpenStax.com questions and answers given on their website or to the College Physics 2e - https://openstax.org/details/books/college-physics-2e.